MASTER

Detection of change points and anomalies in Preferential Attachment Models

Douwes, A.S. (Sander)

Award date:
2024

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics and Computer Science

Detection of change points and anomalies in Preferential Attachment Models

Author:
Sander Douwes - 1021932

Supervisors:
Rui Pires da Silva Castro
Remco van der Hofstad

Eindhoven, June 2024

Abstract

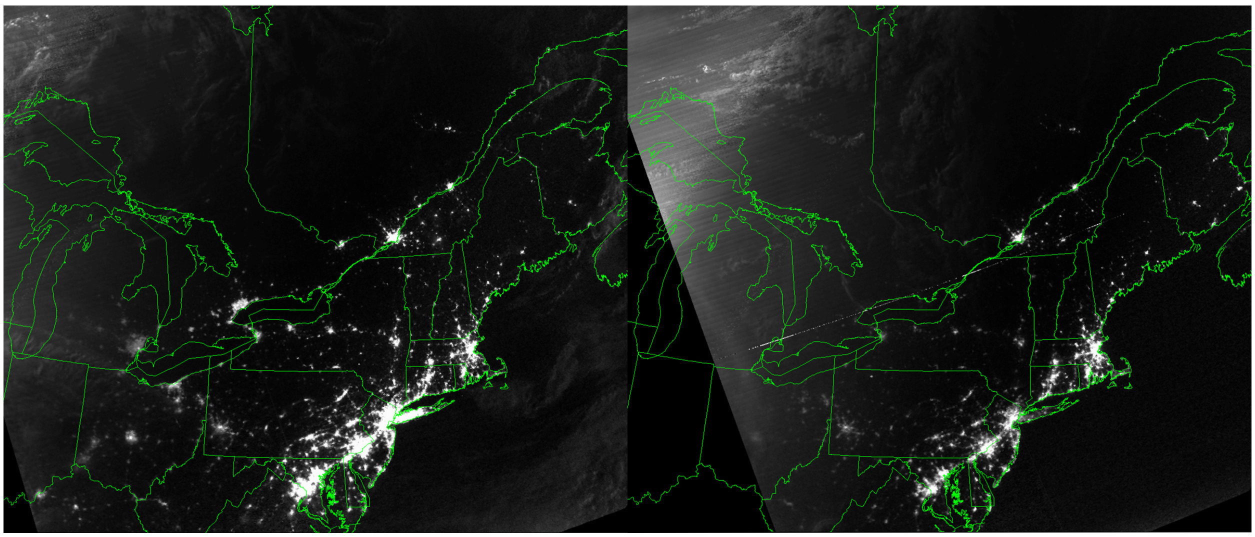
In this thesis we extend key results achieved in earlier research. We analyze Preferential Attachment Models with a change in growth dynamic. Researchers have introduced a statistical test to accurately detect a change in the growth dynamic of a Preferential Attachment Model. In this thesis we prove theoretical results we can use to try to improve the performance of the statistical test. We then introduce an updated statistical test and analyze if the performance is indeed better. Lastly we look into estimating the location of the change in growth dynamic.

Contents

1 Introduction

1.1 Networks

On the evening of August 14 in 2003, the northeastern corridor of the United States was hit with a blackout. The blackout started as a local loss of power, but the local power failures started spreading quickly across the entire power grid. Multiple cities including New York, Boston, Toronto, Detroit and Cleveland were affected. While some cities were able to restore power relatively quickly, most areas affected were left without power for the entire night. This was an example of a network failure. A small defect in one system of the network resulted in a cascading failure across the entire network. Below are two pictures, one taken before the blackout, and one taken after the blackout.



(a) 20 hours before blackout

(b) 7 hours after blackout

Figure 1: NASA satellite images before and after 2003 United States blackout [21]

In Figure 1a you can see a satellite image of northeast United States 20 hours before the blackout, in Figure 1b you can see a satellite image of the same area of the United States 7 hours after the blackout. The satellite images show a clear difference in the intensity of the lights coming from urban areas.

The power grid is one key example, but similar network structures can be found in other real-life phenomena. Examples mentioned in [4] are the 2009-2011 financial meltdown, which began with the burst of the United States housing bubble. This drastically lowered housing prices, resulting in people abandoning their mortgage. This affected the mortgage-backed securities of multiple banks, leading to defaults or government bailouts. Eventually this crisis cascaded into a global financial crisis. A different example from [4] is a man-made cascade. One way scientists try to kill cancer cells in a human body is to trigger a small failure in the cell, eventually leading to its death.

Whether we want to prevent similar cascades in the future or want to use them for our benefit we need to understand the underlying structure of the network we are dealing with. How can a small failure in one system of the network lead to a global crisis? Random graph theory plays a critical role in trying to understand complex networks. Random graphs are used to capture the underlying structure of complex real-life phenomena. By studying random graphs, we can use them to better understand the complex networks present in the world around us. In this thesis we focus on a specific random graph model, the Preferential Attachment Model. The Preferential Attachment Model is a dynamic growth model. This means that over time the model grows bigger in size. We aim to detect drastic changes in the growth dynamic of the model.

1.2 Preferential Attachment Models

Ever since the introduction of the Preferential Attachment Model (PAM in short) in [5], it has been widely used to model many real world phenomena. The world wide web can be modeled using this random graph model [1]. Other interesting applications are the power law relations in the internet [13], in metabolic networks in biology [19] and in scientific collaboration networks [22]. These networks are called a scale-free networks, as the empirical degree distribution closely matches a power law distribution [11]. In a network where the degree distribution closely matches a power law, the proportion of vertices that have k connections is approximately $k^{-\rho}$. The value of ρ typically lies in the interval $(2, 3)$; although other values are possible as well. The way the network is structured gives rise to the so-called small world phenomenon [25], because the distance between 2 arbitrary nodes in the network is small. This means that the number of steps needed to travel from a specific node to another node is low. If two nodes are connected, then it only takes one step. If two nodes are both connected by a third one it will only take two steps. These two properties led the authors of [5] to introduce the PAM as a simple dynamic growth network that exhibits these properties. By using a simple model that captures the main properties of these real-life networks, network scientists have been able to understand the structure of the real-life networks better.

Another useful feature in the Preferential Attachment Model is the dynamic between the number of edges of a vertex, and the likelihood of a new edge making a connection with that vertex. In the PAM, the vertex with the most edges has the largest probability of gaining a new connection from another vertex. This is the 'rich get richer' phenomenon. In our current society, rich people get richer, powerful people get more powerful. This is called the Matthew effect and is studied more closely in [24]. The authors have studied the power law and scaling behaviour in empirical data sets.

A third feature is the growth dynamic of a Preferential Attachment Model. This type of random graph can be used to model an evolving network. In other random graph models, like the Erdős-Rényi Random Graph Model or the Configuration Model, multiple characteristics in the network are determined as soon as the model is started. In an Erdős-Rényi random graph, the number of vertices is always set at the start, whereas the Preferential Attachment Model can handle extra vertices without compromising the underlying structure. In the Configuration Model the degree of a vertex is predetermined. In the Preferential Attachment Model the degree of any vertex can change as soon as a new vertex is introduced in the model.

When the authors of [5] introduced the Preferential Attachment Model, their paper lacked a formal definition. This was later formalized in [11]. The exact construction of the Preferential Attachment Model is also a topic of discussion. The authors of [10] started with a single vertex with m self loops. They also updated the degrees of every vertex during the process that a new vertex is connecting its m edges to existing vertices. In defining the Preferential Attachment Model there are many choices you can make. Do you allow self-loops? If every vertex brings more than one edge, do you update the degree sequence after every attachment? Or do you pick all vertices to attach to and then update the degree sequence? Other versions include a directed Preferential Attachment Model where the vertices have a starting point and an end point. In a directed graph, you are only allowed to traverse an edge from the starting vertex to the end vertex, not the other way. This model is introduced and studied in [9].

Regardless of the way the Preferential Attachment Model is constructed, the most studied aspect of the model is the degree sequence. In the Preferential Attachment Model new vertices attach to existing vertices with degree k proportional to $f(k) = k + \delta(t)$ at time t for some function $\delta(t)$. Then under certain conditions, the expected degree sequence of the PAM follows a specific value p_k, which was first proven by [11] for the case $\delta(t) = 0$. They were able to prove that for $k \leq t^{\frac{1}{15}}$, the expected number of vertices with degree k satisfies the following equation.

$$\mathbb{E}[N_k(t)] = tp_k(1 + o(1)).$$

Then the authors of [18] proved a similar result for $\delta(t) = \delta$ constant, where they were able to prove this without the bound on k. The proof also included cases where $\delta \neq 0$, making this result more general. Other researchers also proved similar results for the expected degree sequence [17].

In this thesis we extend a few key results from [6] and [8]. Instead of a constant growth process, these papers consider a setting where the growth dynamic of the graph undergoes a change. As already mentioned, a new vertex in the model attaches to an already existing vertex with degree k proportional to $f(k) = k + \delta(t)$. If there is no change in growth dynamic, the function $\delta(t)$ remains constant over time ($\delta(t) = \delta$). If there is a change in growth dynamic, then $\delta(t)$ will change after an unknown amount of time.

In this thesis we formulate a statistical test to detect if such a change occurred. We specifically investigate changes that happen late in the growth process of the network. Other researchers have looked into (different variants of) this problem. In [7], the authors analyze the detection of change points for Preferential Attachment Trees. Preferential Attachment Trees are Preferential Attachment Models where every vertex introduced into the graph only gets one edge. In [7] an early change point is considered. The work of [12] also relates to detecting change points in a Preferential Attachment Model. The authors of [12] also consider an early change point in a Preferential Attachment Tree, similar to [7]. The authors of [12] have also considered an extension to the model where not one, but multiple change points were introduced in a Preferential Attachment Tree. They solved this problem by breaking the problem down into multiple single change point detection problems.

This thesis is a direct extension of [6]. We consider exactly the same setting, but we use different test statistics to test the same null hypothesis. In [6] the authors also consider the case where δ is unknown. In this case you have to estimate the value of δ itself. This approach only works if the estimated value is sufficiently 'close' to the actual value. The authors of [14] analyzed this problem when $\delta(t) = \delta$. However, the authors of [6] also introduced an estimator that works under a growth dynamic with a change. By using the assumption that the actual value lies in a known interval ($\delta \in [\delta_{min}, \delta_{max}]$) they were able to give a consistent estimator under both hypotheses that converges in distribution to the actual value. This thesis only considers the first part of [6] where δ is known. The case where δ is unknown is beyond the scope of this thesis.

1.3 Thesis outline

In this thesis we first define the Preferential Attachment Model with or without a change point in Chapter 2. Here we also look at an already existing test from the literature. In Chapter 3 we prove results that will help formulate a new test statistic for detecting a change point. In Chapter 4 we formally introduce the test, then in Chapter 5 the test is compared to the original test from Chapter 2. In Chapter 6 we briefly discuss an estimator for the exact location of the change point. In Chapter 7 we briefly summarize key findings in this thesis, and look for topics of future research.

2 Preliminaries

In this chapter we first define the Preferential Attachment Model. In Section 2.1 we define the base model, then we define a change point and how it alters the model. We also discuss the data that is available for use in a statistical test and the data that is not available. In Section 2.2 we introduce a statistical test from the literature. We describe how the test works, then we discuss how we aim to improve the test from literature in this thesis. Before we introduce the Preferential Attachment Model, we need to define some notation. For a sequence of random variables X_n, we write $X_n \overset{D}{\to} X$ if X_n converges in distribution to X, and we write $X_n \overset{\mathbb{P}}{\to} X$ if X_n converges in probability to X.

2.1 Model description

Now we are ready to introduce the Preferential Attachment Model. The model is defined as follows. Let $G_n = (V_n, E_n)$ be an undirected random graph without self loops. Here $V_n = \{v_0, \dots, v_n\}$ is the vertex set and $E_n \subseteq \{(i, j) : i, j \in V_n; i \neq j\}$ denotes a random edge set. We first describe how the random graph is constructed. We start with an initial graph G_1, which has two vertices. These two vertices are connected to each other with m edges. To get G_t from G_{t-1}, we consider the graph $G_{t,0}$ as G_{t-1} with an isolated vertex v_t. The isolated vertex gets m edges to connect to vertices in $G_{t,0}$. This results in m intermediate graphs $G_{t,1}, \dots, G_{t,m}$. We define $D_v(G_{t,i})$ as the degree of vertex v in graph $G_{t,i}$. Vertex v is picked with conditional probability

$$\mathbb{P}(v_{t,i} = v | G_{t,i-1}) = \frac{D_v(G_{t,i-1}) + \delta(t)}{\sum_{j=0}^{t-1} \big(D_j(G_{t,i-1}) + \delta(t)\big)}. \tag{1}$$

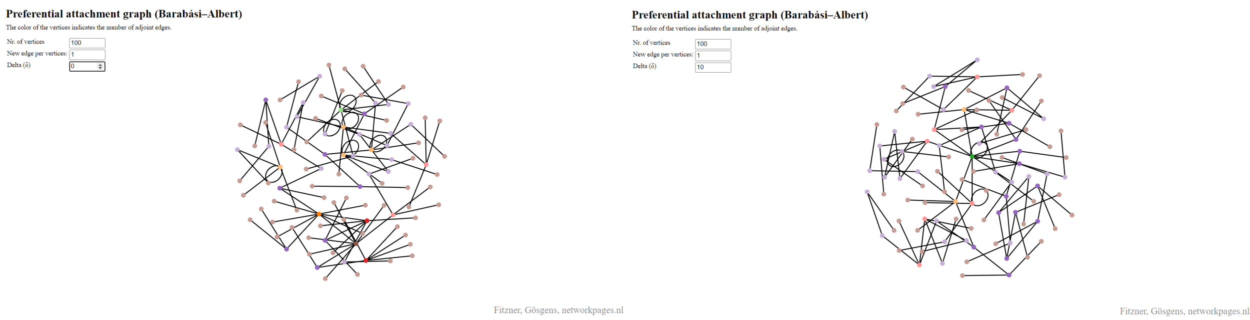
After picking the first vertex to connect to, we get the updated graph $G_{t,1}$. Then we pick another vertex using Equation (1), but with $G_{t,1}$ instead of $G_{t,0}$. This process repeats until we get $G_{t,m}$. At this point we set $G_{t,m} = G_t$. This process repeats until we obtain the final graph G_n. We see that graph G_n is made using two parameters: $m \in \mathbb{N}$ and $\delta : \mathbb{N} \to (-m, \infty)$.

If we look at Equation (1), we see that the probability of picking vertex v is proportional to the degree of v plus the function $\delta(t)$. We give a short intuitive explanation of what the function $\delta(t)$ does. The function $\delta(t)$ can increase or decrease the chance that a vertex gets chosen. If $\delta(t) > 0$, vertices with a low degree have a higher probability of receiving a new connection and vertices with a high degree have a lower probability of receiving a new connection. If we set $\delta(t) < 0$, the opposite happens. Vertices with high degree get an increased chance of receiving a new connection, vertices with low degree see their chances decrease. While this model is flexible, the condition $\delta(t) > -m$ has to be met, otherwise Equation (1) might not be a valid probability.

We consider the Affine PAM model where the function $\delta(t)$ is equal to a constant δ_0. The exact value of δ_0 will determine the structure of the final graph G_n. Using [15] we can visualize two graphs made using a different value for $\delta(t)$. Note that the creators of [15] have made their model with self-loops allowed. The graph in Figure 2a has been made using $\delta(t) = 0$, the graph in Figure 2b has been made using $\delta(t) = 10$. By looking at the final result of the process, it might not be clear that $\delta(t)$ is different in the two instances. This information is hidden in the properties of the graph. In [18], all of these properties are explained and proven. We look at one specific property. We are interested in the number of vertices that have a certain degree $k \geq m$. To analyze this quantity, we first state some results from [18]. We first introduce $p_k(\delta_0)$ as

$$p_k(\delta_0) = \left(2 + \frac{\delta_0}{m}\right) \frac{\Gamma(k + \delta_0)\Gamma(m + 2 + \delta_0 + \frac{\delta_0}{m})}{\Gamma(k + 3 + \delta_0 + \frac{\delta_0}{m})\Gamma(m + \delta_0)}.$$

In [18] it is proven that $(p_k(\delta_0))_{k \geq m}$ arises as the limiting degree distribution for the Preferential Attachment Model, as introduced in this chapter. Let $N_k(n)$ be the number of vertices with degree k at time n. Then



(a) $\delta(t) = 0$.

(b) $\delta(t) = 10$.

Figure 2: Realizations of two random graphs with different values for $\delta(t)$.

by Proposition 8.4 and 8.7 in [18], for $C > 0$ and $n \geq 1$,

$$\mathbb{P}(\max_k |N_k(n) - np_k(\delta_0)| \geq C(1 + \sqrt{n \log(n)})) = o(1). \tag{2}$$

Equation (2) implies that $\frac{N_k(n)}{n} \xrightarrow{\mathbb{P}} p_k(\delta_0)$. This means that the proportion of vertices with fixed degree k will converge in probability to $p_k(\delta_0)$.

These results show us that we know the behaviour of the number of vertices with degree k. We know that this behaviour is dependent on (1), and specifically on the function $\delta(t)$. To continue, we consider the same model, but now we allow $\delta(t)$ to change over time. We specifically look into models where the function $\delta(t)$ is a step function. Let $c > 0$, $\gamma \in (0, 1)$ and τ_n be a point in time of the form

$$\tau_n = n - cn^\gamma. \tag{3}$$

Then in the construction of a graph with a change point, the (step) function $\delta(t)$ is defined as

$$\delta(t) = \mathbb{1}\{t \leq \tau_n\}\delta_0 + \mathbb{1}\{t > \tau_n\}\delta_1. \tag{4}$$

We refer to τ_n as the change point, the point where the function $\delta(t)$ changes value. This means that after the change point, the probability that a new vertex connects with an existing vertex changes.

In this thesis the main question we want to answer is: has $\delta(t)$ been changed during the creation of the graph? We aim to answer this question with yes or no, based on available data. In order to check if the function $\delta(t)$ has been changed during the creation of a graph, we formulate a statistical hypothesis test. Under the null model, defined as H_0, we assume that $\delta(t) = \delta_0$ constant. Under the alternative hypothesis (H_1) we assume that $\delta(t)$ is a step function as defined in Equation (4). We also assume under H_1 that $\delta_1 \neq \delta_0$ and $\tau_n \in \mathbb{N}$ with $\tau_n < n$. We aim to show that we can distinguish between the two cases with high accuracy. On the one hand, we want to be confident in the test. If the null hypothesis is true, our test should reject the null hypothesis with low probability. On the other hand, we want the test to have a high power. If the alternative hypothesis is true, our test should reject the null hypothesis with high probability. There is always a trade off between the power and confidence of a statistical test. We go into more detail on the power, confidence and the trade-off between them in Chapter 4.

The only data that we get to distinguish the two hypotheses is the final graph G_n without any labels, similar to the graphs generated in Figures 2a and 2b. Because we know how G_n is constructed we can extract some information from the graph. The final version of the graph has $n+1$ vertices and $n \cdot m$ edges, so we can infer m. We can also calculate the degree sequence of the graph based only on G_n. We assume that δ_0 is known. The other parameters δ_1, c and γ are unknown. We give a short overview of the data that is available to us, and the data that is not available to us.

- Graph G_n is available (without labels).

- Number of vertices in G_n is known.
- Number of edges introduced per vertex (m) is known.
- Degree sequence of G_n is known.
- The parameter δ_0 is assumed to be known.
- Intermediate graphs used in the construction of G_n are unavailable.
- The parameter δ_1 is unknown.
- The parameter c is unknown.
- The parameter γ is unknown.

2.2 Literature review

The authors of [6] and [8] already proved that you can accurately test for a change point based on the information given. Using only the number of vertices with degree m ($N_m(n)$), they were able to distinguish between the null hypothesis and the alternative hypothesis with high accuracy under specific conditions. We now introduce this test from the literature. To be able to introduce the asymptotically calibrated minimal degree test from [6], we first define the test statistic. We define $T(G_n) = N_m(n) - np_m(\delta_0)$. Second, define

$$w(\delta_0, m) = \frac{m^2(m+\delta_0)(m+\delta_0+1)(2m+\delta_0)}{(2m(m+\delta_0+1)+\delta_0)(m^2+m\delta_0+2m+\delta_0)^2}.$$

Then for $n \to \infty$ and $\delta(t) = \delta_0$, the authors of [6] proved that

$$\frac{T(G_n)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}\left(0, w(\delta_0, m)\right).$$

This result shows that for $\delta(t) = \delta_0$ fixed we know the asymptotic behaviour of the test statistic $T(G_n)$. Now we have the information needed to introduce the test from [6]. We assume there is no change point under the null hypothesis (H_0), and under the alternative hypothesis (H_1) there is a change point of the form described in (3), and $\delta(t)$ is defined as (4). Let z_α denote the right quantile function of a standard normal distribution at significance level α. The asymptotically calibrated minimal degree test is formulated as

$$\psi_{cal}(G_n) = \mathbb{1}\{|T(G_n)| \geq \sqrt{nw(\delta_0, m)}z_{\alpha/2}\}. \tag{5}$$

If $\psi_{cal}(G_n) = 1$ we reject the null hypothesis. If $\psi_{cal}(G_n) = 0$ we fail to reject the null hypothesis. The test defined in (5) is based on the fact that $T(G_n)$ is approximately normally distributed with mean equal to zero and variance equal to $nw(\delta_0, m)$ under the null model. Let $Z \sim \mathcal{N}(0, 1)$ be a standard normal random variable. Then under the null model,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(-\sqrt{nw(\delta_0, m)}z_{\alpha/2} \leq T(G_n) \leq \sqrt{nw(\delta_0, m)}z_{\alpha/2}) \\
&= \lim_{n\to\infty} \mathbb{P}\left(-z_{\alpha/2} \leq \frac{T(G_n)}{\sqrt{nw(\delta_0, m)}} \leq z_{\alpha/2}\right) \\
&= \mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha.
\end{aligned}$$

This shows that the confidence of the test under significance level α is equal to $1 - \alpha$. By increasing or decreasing α we increase or decrease the confidence of the test. However, under the alternative model $N_m(n)$

will not necessarily converge to $p_m(\delta_0)$. In [6] it is shown that if $\delta(t) = \mathbb{1}\{t \leq \tau_n\}\delta_0 + \mathbb{1}\{t > \tau_n\}\delta_1$ and $n \to \infty$,

$$\mathbb{E}[T(G_n)] = \mathbb{E}[N_m(n)] - np_m(\delta_0) = (1 + o(1))cn^\gamma \frac{\delta_0 - \delta_1}{(2 + \frac{\delta_1}{m})(m + \delta_0 + 2 + \frac{\delta_0}{m})}. \tag{6}$$

This proves that for $n \to \infty$, $c > 0$ and $\gamma \geq \frac{1}{2}$ the expected value of $T(G_n)$ is not equal to zero when the alternative hypothesis is true. This is the power behind the test. Under the null model, the statistic $T(G_n)$ will be close to a normal random variable with mean zero. The authors of [6] showed that if a change point is present, $T(G_n)$ will be close to a normal random variable with mean equal to (6). This allows the test to detect the existence of a change point. Note that Equation (6) only holds when $\delta(t)$ is a step function. For different functions different results will hold.

The test defined in Equation (5) uses the number of vertices with minimal degree ($N_m(n)$) to test for the existence of a change point. We already know that the final version of the graph (G_n) has more information available. We try to improve the test as introduced in (5) by including more information. We specifically focus on adding one extra piece of information, the number of vertices with degree $m + 1$, namely $N_{m+1}(n)$. Before we can use this piece of information we need to prove key results first. This will be done in Chapter 3.

3 Characteristics of degree counts

The presence of a change point in a Preferential Attachment Model can be seen in the empirical degree distribution. The test statistics that we consider in this thesis are based on counts on the number of vertices with a certain degree. In this chapter we provide a characterization of these counts when a change point is present. We prove two key results. In Section 3.1 we first prove a result for the expected value of the degree count under the alternative hypothesis. Section 3.1.1 will focus on the theoretical analysis, and in Section 3.1.2 we will provide a simulation to check the theoretical analysis. In Section 3.2 we calculate the variance and covariance of degree counts. Here we again do a theoretical calculation in Section 3.2.1, and then provide a simulation in Section 3.2.2. This will give us the necessary tools to introduce a new test statistic.

3.1 Expected value of degree counts under alternative hypothesis

To be able to use the extra information available in the final graph G_n, we need to know how the Preferential Attachment Model behaves under the alternative hypothesis H_1. In [6], this behaviour is analyzed for the minimal degree m in order to formulate the test as introduced in (5). Under H_1, the expected value of $N_m(n) - np_m(\delta_0)$ as $n \to \infty$ is equal to

$$\mathbb{E}[N_m(n)] - np_m(\delta_0) = (1 + o(1))cn^{\gamma}\frac{\delta_0 - \delta_1}{(2 + \frac{\delta_1}{m})(m + \delta_0 + 2 + \frac{\delta_0}{m})}.$$

In the final graph we not only know the number of vertices with minimal degree, we know the number of vertices with degree k for every $k \geq m$. To use this additional information, we analyze the behaviour of vertices with general degree k. We prove a proposition that characterizes the degree count when a change point is present for fixed $k > m$ in Section 3.1.1. Then in Section 3.1.2 we provide a simulation of the proposition.

3.1.1 Theoretical analysis

We first state the proposition that we are going to prove. The authors of [6] proved equation (6) for the expected number of vertices with degree m, we do the same for vertices with general degree k. Recall that $p_k(\delta_0)$ arises as the limiting degree distribution for $\frac{N_k(n)}{n}$. We also introduce $l \in \{0, 1\}$ to indicate a calculation under both hypotheses. If a certain calculation is done under the null hypothesis, we fill in $l = 0$. If a result is calculated under the alternative hypothesis we fill in $l = 1$. As an example, $\mathbb{E}_l[\cdot]$ is the expected value under both hypotheses, $\mathbb{E}_0[\cdot]$ and $\mathbb{E}_1[\cdot]$ are the expected values under the null and alternative hypothesis respectively. Now we can introduce Proposition 1.

Proposition 1. Let $\gamma < 1$ and $k = o(n^{1-\gamma})$. Then for $n \to \infty$,

$$\mathbb{E}_1[N_k(n)] - np_k(\delta_0) = cn^{\gamma}(\delta_0 - \delta_1)\left(\frac{(2m - k)}{(2 + \frac{\delta_0}{m})(2m + \delta_1)}p_k(\delta_0) - \frac{(2m - (k-1))}{(2 + \frac{\delta_0}{m})(2m + \delta_1)}p_{k-1}(\delta_0)\right) + o(n^{\gamma}). \quad (7)$$

Proposition 1 extends the result obtained by [6] to general degree k. However, the proposition states that the constraint $k = o(n^{1-\gamma})$ has to be met. This is a consequence of the rich get richer phenomenon present in Preferential Attachment Models. If k becomes too large, the vertex with degree k will receive preferential attachment. In this case Proposition 1 will not hold. We now prove Proposition 1.

Proof of Proposition 1. In Proposition 1 we see that the final result is order $O(n^{\gamma})$. We first rewrite the equation into three distinct terms which can be solved individually. These terms require further analysis, where we make a distinction among vertices arriving before the change point τ_n and vertices arriving after.

Then we show that a lot of terms in the analysis will have an order that is strictly lower than $O(n^\gamma)$ when $\gamma < 1$. When $n \to \infty$, the lower order terms become insignificant. We start by splitting the equation into three distinct terms. We know that under H_1,

$$\mathbb{E}_1[N_k(n)] - np_k(\delta_0) = \big(\mathbb{E}_0[N_k(n)] - np_k(\delta_0)\big) + \big(\mathbb{E}_1[N_k(n)] - \mathbb{E}_1[N_k(\tau_n)]\big) - \big(\mathbb{E}_0[N_k(n)] - \mathbb{E}_0[N_k(\tau_n)]\big). \tag{8}$$

Note that before the change point τ_n the expected number of vertices with degree k is identical for both models. Therefore $\mathbb{E}_1[N_k(\tau_n)] = \mathbb{E}_0[N_k(\tau_n)]$, and Equation (8) is correct. The first term of this equation is bounded as follows. By Proposition 8.7 of [18] there exists a $C_0 = C_0(\delta_0, k)$ such that

$$|\mathbb{E}_0[N_k(n)] - np_k(\delta_0)| \le C_0 = O(1).$$

The bound above takes care of the first term of (7). Now we have two more terms to go. Notice that these terms are almost identical. The only difference between them is that the first term is calculated under the alternative hypothesis H_1, and the second term is calculated under the null hypothesis H_0. Therefore we analyse these terms together. The first result holds for both hypotheses. Let $l \in \{0, 1\}$, then

$$\mathbb{E}_l[N_k(n)] - \mathbb{E}_l[N_k(\tau_n)] = \sum_{v \in [\tau_n]} \Big(\mathbb{P}_l(D_v(n) = k) - \mathbb{P}_l(D_v(\tau_n) = k)\Big) + \sum_{v \in [n] \setminus [\tau_n]} \mathbb{P}_l(D_v(n) = k). \tag{9}$$

We decompose Equation (9) into two distinct sums. The first sum captures the contribution of the vertices already present at the change point. The second sum calculates the contribution of vertices arriving after the change point. For the case $k = m$, the authors of [6] already analyzed how these terms behave. We consider the case $k > m$. We first look at the last sum in Equation (9). Let $k > m$ be fixed and $v \in [n] \setminus [\tau_n]$. Then

$$\begin{aligned}
\mathbb{P}_l(D_v(n) = k) &\le \mathbb{P}_l(D_v(n) > m) \\
&= \mathbb{P}_l(\exists t \in [\tau_n, n] : D_v(t) > m) \\
&\le m(n - \tau_n) \frac{m + \delta_l}{(2m + \delta_l)\tau_n - 2m} \\
&= m(cn^\gamma) \frac{m + \delta_l}{(2m + \delta_l)(n - cn^\gamma) - 2m} \\
&= O(n^{\gamma - 1}).
\end{aligned}$$

The event $D_v(t) > m$ can only happen if there is at least one vertex v' introduced after the change point that connects with vertex v. The odds of this happening is at most $(m + \delta_l)/((2m + \delta_l)\tau_n - 2m)$ and there are at most $m(n - \tau_n)$ edges that can connect with v. Using this result we get

$$\begin{aligned}
\sum_{v \in [n] \setminus [\tau_n]} \mathbb{P}_l(D_v(n) = k) &\le (n - \tau_n) O(n^{\gamma - 1}) = cn^\gamma O(n^{\gamma - 1}) \\
&= O(n^{2\gamma - 1}).
\end{aligned}$$

As we have seen, Equation (7) has order $O(n^\gamma)$. For all $\gamma < 1$ we have that $O(n^{2\gamma - 1}) < O(n^\gamma)$. This implies that the vertices that arrive after the change point have a smaller order than the vertices arriving before the change point. Because the order of the contribution of the vertices after the change point is lower, their contribution will be insignificant when $n \to \infty$. In other words, if the number of vertices is large enough, the contribution of vertices $v' \in [n] \setminus [\tau_n]$ will become so small they will not be noticed anymore. This means for the analysis this term will be ignored. An important exception is the case $\gamma = 1$, then both terms have the same order. We assume $\gamma < 1$ for the remainder of the derivation. For the other term, we condition on the degree of v at the change point $(D_v(\tau_n))$ to obtain

$$\begin{aligned}
&\mathbb{P}_l(D_v(n) = k) - \mathbb{P}_l(D_v(\tau_n) = k) \\
&= \sum_{j=m}^{k} \Big(\mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j) \mathbb{P}_l(D_v(\tau_n) = j)
\end{aligned}$$

$$
\begin{aligned}
&\quad - \mathbb{P}_l(D_v(\tau_n) = k | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j)\Big) \\
&= \sum_{j=m}^{k} \left(\mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j) - \mathbb{P}_l(D_v(\tau_n) = k | D_v(\tau_n) = j)\right) \mathbb{P}_l(D_v(\tau_n) = j) \\
&= \left(\mathbb{P}_l(D_v(n) - D_v(\tau_n) = 0 | D_v(\tau_n) = k) - 1\right) \mathbb{P}_l(D_v(\tau_n) = k) \\
&\quad + \sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j).
\end{aligned}
$$

To investigate these terms, we recall Lemma 7.2 from [6]:

Lemma 7.2. Let $v \in [\tau_n], \gamma < 1$, $m \leq k = o(n^{1-\gamma})$ and $l \in \{0, 1\}$. Then as $n \to \infty$,

$$
\mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k) = (1 + o(1))cn^{\gamma-1}m\frac{k + \delta_l}{2m + \delta_l}. \tag{10}
$$

To calculate the first term, we use the complement of (10). The lemma is for the event $D_v(n) - D_v(\tau_n) > 0$, we take the event $D_v(n) - D_v(\tau_n) = 0$. This can be obtained by taking $1 - \mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k)$. Using this we obtain

$$
\begin{aligned}
&\mathbb{P}_l(D_v(n) = k) - \mathbb{P}_l(D_v(\tau_n) = k) \\
&= \left(-(1 + o(1))cn^{\gamma-1}m\frac{k + \delta_l}{2m + \delta_l}\right) \mathbb{P}_l(D_v(\tau_n) = k) + \sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j).
\end{aligned} \tag{11}
$$

In order to calculate the last term, we introduce Proposition 2:

Proposition 2. Let $\gamma < 1$ and $l \in \{0, 1\}$. Then as $n \to \infty$,

$$
\begin{aligned}
&\sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j) \\
&= \mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k - 1)\mathbb{P}_l(D_v(\tau_n) = k - 1) + O(n^{2(\gamma-1)}).
\end{aligned} \tag{12}
$$

Proposition 2 states that the main contribution towards the amount of vertices that get degree k in the final version of the graph comes from vertices that already have degree k and receive no attachments, or from vertices with degree $k - 1$ at the change point and receive one attachment. If a vertex v has degree $k - 2$ or lower at time τ_n, the probability that $D_v(n) = k$ is very low. The contribution from vertices with degree equal to $k - 2$ or less will vanish when $n \to \infty$.

Proof of Proposition 2. To prove Proposition 2, we look at the difference between the first two terms:

$$
\begin{aligned}
&\left(\sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j)\right) \\
&\qquad - \mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k - 1)\mathbb{P}_l(D_v(\tau_n) = k - 1) \\
&= \left(\sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) - D_v(\tau_n) = k - j | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j)\right) \\
&\qquad - \mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k - 1)\mathbb{P}_l(D_v(\tau_n) = k - 1).
\end{aligned}
$$

We notice that the right hand side of the proposition is slightly larger than the $j = k - 1$ term on the left hand side. This gives us the upper bound

$$
\left(\sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) - D_v(\tau_n) = k - j | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j)\right)
$$

$$
\begin{aligned}
&- \mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k-1)\mathbb{P}_l(D_v(\tau_n) = k-1) \\
&\le \sum_{j=m}^{k-2} \mathbb{P}_l(D_v(n) - D_v(\tau_n) = k-j | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j) \\
&\le \binom{m(n-\tau_n)}{2}\Big(\frac{k-2+\delta_l}{(2m+\delta_l)\tau_n - 2m} \times \frac{k-1+\delta_l}{(2m+\delta_l)\tau_n - 2m}\Big) \\
&= O(n^{2\gamma-2}) = O(n^{2(\gamma-1)}).
\end{aligned}
$$

This bound shows that the probability that a vertex v with degree $k-2$ or less reaches degree k in the final version of the graph is order $O(n^{2(\gamma-1)})$. We again note that the other terms in the derivation have order $O(n^{\gamma})$. This means that for all $\gamma < 1$, $O(n^{2(\gamma-1)}) < O(n^{\gamma})$. We can use this fact to substitute the second term in for the first term in the remainder of the proof.

We now continue the proof of Proposition 1. By Proposition 2, the last term of (11) can be computed in the following way. We first substitute the second term of (12) in for the sum, and then we immediately apply Lemma 7.2 (10) again to get

$$
\begin{aligned}
&\mathbb{P}_l(D_v(n) = k) - \mathbb{P}_l(D_v(\tau_n) = k) \\
&= \left(-(1+o(1))cn^{\gamma-1}m\frac{k+\delta_l}{2m+\delta_l}\right)\mathbb{P}_l(D_v(\tau_n) = k) \\
&\quad + \sum_{j=m}^{k-1} \mathbb{P}_l(D_v(n) = k | D_v(\tau_n) = j)\mathbb{P}_l(D_v(\tau_n) = j) \\
&= \left(-(1+o(1))cn^{\gamma-1}m\frac{k+\delta_l}{2m+\delta_l}\right)\mathbb{P}_l(D_v(\tau_n) = k) \\
&\quad + \mathbb{P}_l(D_v(n) - D_v(\tau_n) > 0 | D_v(\tau_n) = k-1)\mathbb{P}_l(D_v(\tau_n) = k-1) + O(n^{2(\gamma-1)}) \\
&= \left(-(1+o(1))cn^{\gamma-1}m\frac{k+\delta_l}{2m+\delta_l}\right)\mathbb{P}_l(D_v(\tau_n) = k) \\
&\quad + \left((1+o(1))cn^{\gamma-1}m\frac{k-1+\delta_l}{2m+\delta_l}\right)\mathbb{P}_l(D_v(\tau_n) = k-1) + O(n^{2(\gamma-1)}).
\end{aligned}
$$

Now we look back at Equation (8). Here the terms are calculated under a specific hypothesis. We use the previous results, but now we fill in either $l = 0$ or $l = 1$. This results in

$$
\begin{aligned}
&\mathbb{E}_1[N_k(n)] - np_k(\delta_0) \\
&= \mathbb{E}_0[N_k(n)] - np_k(\delta_0) + \mathbb{E}_1[N_k(n)] - \mathbb{E}_1[N_k(\tau_n)] - (\mathbb{E}_0[N_k(n)] - \mathbb{E}_0[N_k(\tau_n)]) \\
&= O(1) + \mathbb{E}_1[N_k(n)] - \mathbb{E}_1[N_k(\tau_n)] - (\mathbb{E}_0[N_k(n)] - \mathbb{E}_0[N_k(\tau_n)]) \\
&= O(1) + O(n^{2\gamma-1}) + \sum_{v\in[\tau_n]} \Big(-(1+o(1))cn^{\gamma-1}m\frac{k+\delta_1}{2m+\delta_1}\mathbb{P}_1(D_v(\tau_n) = k) \\
&\qquad + (1+o(1))cn^{\gamma-1}m\frac{k-1+\delta_1}{2m+\delta_1}\mathbb{P}_1(D_v(\tau_n) = k-1) \\
&\qquad + (1+o(1))cn^{\gamma-1}m\frac{k+\delta_0}{2m+\delta_0}\mathbb{P}_0(D_v(\tau_n) = k) \\
&\qquad - (1+o(1))cn^{\gamma-1}m\frac{k-1+\delta_0}{2m+\delta_0}\mathbb{P}_0(D_v(\tau_n) = k-1) + O(n^{2(\gamma-1)})\Big).
\end{aligned}
$$

We notice that before the change point the probability that a vertex has certain degree k is identical under the null and alternative hypothesis. We can combine the terms corresponding with $\mathbb{P}_1(D_v(\tau_n) = k)$ and

$\mathbb{P}_0(D_v(\tau_n) = k)$ together. We can do the same with the terms corresponding to $\mathbb{P}_1(D_v(\tau_n) = k-1)$ and $\mathbb{P}_0(D_v(\tau_n) = k-1)$ to obtain

$$\begin{aligned}
&\mathbb{E}_1[N_k(n)] - np_k(\delta_0) \\
&= O(1) + O(n^{2\gamma-1}) + \sum_{v\in[\tau_n]} \Bigg((1+o(1))cn^{\gamma-1}m(\delta_0-\delta_1)\frac{2m-k}{(2m+\delta_1)(2m+\delta_0)}\mathbb{P}_0(D_v(\tau_n)=k) \\
&\quad - (1+o(1))cn^{\gamma-1}m(\delta_0-\delta_1)\frac{2m-(k-1)}{(2m+\delta_1)(2m+\delta_0)}\mathbb{P}_0(D_v(\tau_n)=k-1) + O(n^{2(\gamma-1)})\Bigg).
\end{aligned}$$

The proof is almost complete. We only need to rewrite $\mathbb{P}_0(D_v(\tau_n) = k)$ and $\mathbb{P}_0(D_v(\tau_n) = k-1)$. To rewrite these terms, we use the following equation. For $k \geq m$ and $\gamma < 1$,

$$\begin{aligned}
&\sum_{v\in[\tau_n]} \mathbb{P}_0(D_v(\tau_n)=k) \\
&= \sum_{v\in[\tau_n]} \mathbb{E}_0[\mathbb{1}\{D_v(\tau_n)=k\}] = \mathbb{E}_0[N_k(\tau_n)] = (\tau_n p_k(\delta_0) + o(n)) = np_k(\delta_0) + o(n) + O(n^\gamma). \qquad (13)
\end{aligned}$$

Equation (13) states that if we sum $\mathbb{P}(D_v(\tau_n) = k)$ over all vertices that arrived before the change point, we get the expected number of vertices that have degree k at the time of the change point τ_n. We can then rewrite this into $p_k(\delta_0)$ using the relationship between $N_k(\tau_n)$ and $\tau_n p_k(\delta_0)$ as introduced in (2). One more important thing to note is that the order term changes as well. When summed over $v \in [\tau_n]$, the order term changes to

$$\sum_{v\in[\tau_n]} O(n^{2(\gamma-1)}) = (n - cn^\gamma)O(n^{2(\gamma-1)}) = O(n^{2\gamma-1}).$$

Note that this order is still smaller than order $O(n^\gamma)$ when $\gamma < 1$. In fact, it has exactly the same order as the contribution of vertices arriving after the change point. We plug (13) into the equation and obtain

$$\begin{aligned}
&\mathbb{E}_1[N_k(n)] - np_k(\delta_0) \\
&= O(1) + \Bigg((1+o(1))cn^{\gamma-1}m(\delta_0-\delta_1)\frac{2m-k}{(2m+\delta_1)(2m+\delta_0)}\mathbb{E}_0[N_k(\tau_n)] \\
&\quad - (1+o(1))cn^{\gamma-1}m(\delta_0-\delta_1)\frac{2m-(k-1)}{(2m+\delta_1)(2m+\delta_0)}\mathbb{E}_0[N_{k-1}(\tau_n)] + O(n^{2\gamma-1})\Bigg) \\
&= O(1) + \Bigg((1+o(1))cn^{\gamma-1}m(\delta_0-\delta_1)\frac{2m-k}{(2m+\delta_1)(2m+\delta_0)}\left(np_k(\delta_0) + o(n) + O(n^\gamma)\right) \\
&\quad - (1+o(1))cn^{\gamma-1}m(\delta_0-\delta_1)\frac{2m-(k-1)}{(2m+\delta_1)(2m+\delta_0)}\left(np_{k-1}(\delta_0) + o(n) + O(n^\gamma)\right)\Bigg) + O(n^{2\gamma-1}). \\
&= O(1) + \Bigg((1+o(1))cn^{\gamma}m(\delta_0-\delta_1)\frac{2m-k}{(2m+\delta_1)(2m+\delta_0)}\left(p_k(\delta_0) + \frac{o(n)}{n} + O(n^{\gamma-1})\right) \\
&\quad - (1+o(1))cn^{\gamma}m(\delta_0-\delta_1)\frac{2m-(k-1)}{(2m+\delta_1)(2m+\delta_0)}\left(p_{k-1}(\delta_0) + \frac{o(n)}{n} + O(n^{\gamma-1})\right)\Bigg) + O(n^{2\gamma-1}).
\end{aligned}$$

To obtain the final result, we combine both equations. We first look at the remaining order terms. The first term is an $\frac{o(n)}{n}$ term. By definition of the small o notation, this terms goes to zero as $n \to \infty$. The second term is an $O(n^{\gamma-1})$ term. By assumption $\gamma < 1$, so $O(n^{\gamma-1}) = o(1)$ and will also go to zero. We also have an order $O(n^{2\gamma-1})$ term and an order $O(1)$ term. As we already explained, these terms always have lower order when $\gamma < 1$. Because all of these terms are lower order than the leading $O(n^\gamma)$ term, the contribution will become insignificant when $n \to \infty$. However, there is an exception. Under specific circumstances, the multiplicative factor involving $p_k(\delta_0)$ and $p_{k-1}(\delta_0)$ can be zero. In this case the leading order term becomes zero, which means the lower order terms become leading. We capture this behaviour by transforming the lower order terms into one order term: $o(n^\gamma)$. Combining the leading order terms and solving the remaining order terms yields

$$\begin{aligned}
&\mathbb{E}_1[N_k(n)] - np_k(\delta_0) \\
&= O(1) + (1+o(1))cn^\gamma m(\delta_0-\delta_1)\frac{(2m-k)}{(2m+\delta_0)(2m+\delta_1)}p_k(\delta_0) \\
&\quad - (1+o(1))cn^\gamma m(\delta_0-\delta_1)\frac{2m-(k-1)}{(2m+\delta_1)(2m+\delta_0)}p_{k-1}(\delta_0) + O(n^{2\gamma-1}) \\
&= cn^\gamma m(\delta_0-\delta_1)\left(\frac{(2m-k)}{(2m+\delta_0)(2m+\delta_1)}p_k(\delta_0) - \frac{(2m-(k-1))}{(2m+\delta_0)(2m+\delta_1)}p_{k-1}(\delta_0)\right) + o(n^\gamma) \\
&= cn^\gamma(\delta_0-\delta_1)\left(\frac{(2m-k)}{(2+\frac{\delta_0}{m})(2m+\delta_1)}p_k(\delta_0) - \frac{(2m-(k-1))}{(2+\frac{\delta_0}{m})(2m+\delta_1)}p_{k-1}(\delta_0)\right) + o(n^\gamma).
\end{aligned}$$

This completes the proof □.

In Section 3.1.2 we use simulation to confirm Proposition 1.

3.1.2 Simulation results

We check Proposition 1 by simulation for different scenarios. Using the algorithm in Appendix B, we simulate 2000 Preferential Attachment Models with a change point. After simulation we have 2000 values for $N_k(n)$ for $m \leq k < 20$. We subtract the empirical value under the null ($np_k(\delta_0)$) for every estimate. Then we take the average value for every k to get a good estimate of the simulated value. We compare this average with Proposition 1 (7). We start with the values $\delta_0 = 0$, $\delta_1 = 1$, $m = 1$, $n = 10000$, $c = 1$, $\gamma = \frac{1}{2}$. We define $f(k)$ as

$$f(k) = cn^\gamma(\delta_0-\delta_1)\left(\frac{(2m-k)}{(2+\frac{\delta_0}{m})(2m+\delta_1)}p_k(\delta_0) - \frac{(2m-(k-1))}{(2+\frac{\delta_0}{m})(2m+\delta_1)}p_{k-1}(\delta_0)\right).$$

We plot $f(k)$ below.

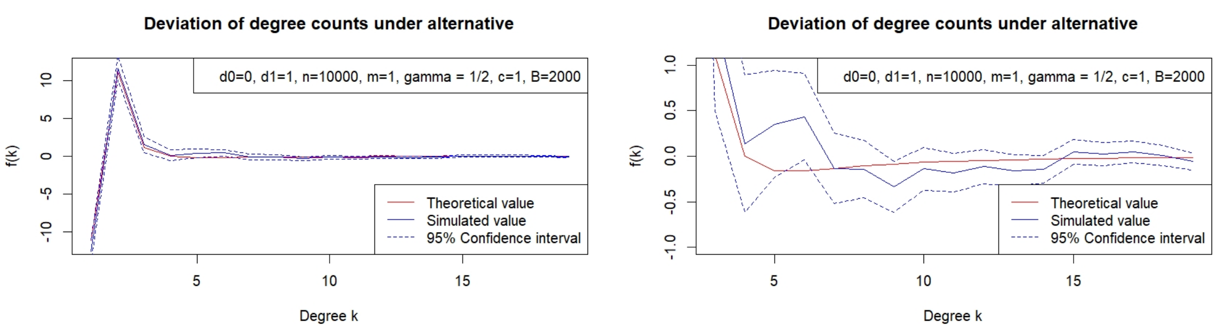


Figure 3: Deviation of the degree counts with $\delta_0 = 0, \delta_1 = 1, m = 1, n = 10000, c = 1, \gamma = \frac{1}{2}$.

In Figure 3 we see on a macroscopic scale that the simulation and the theoretical value coincide well. When zoomed in on the right side, we see that the theoretical value does not exactly match the simulated value. As Equation (7) was proven for $n \to \infty$, we increase the number of vertices to 50000 and run the same simulation again.

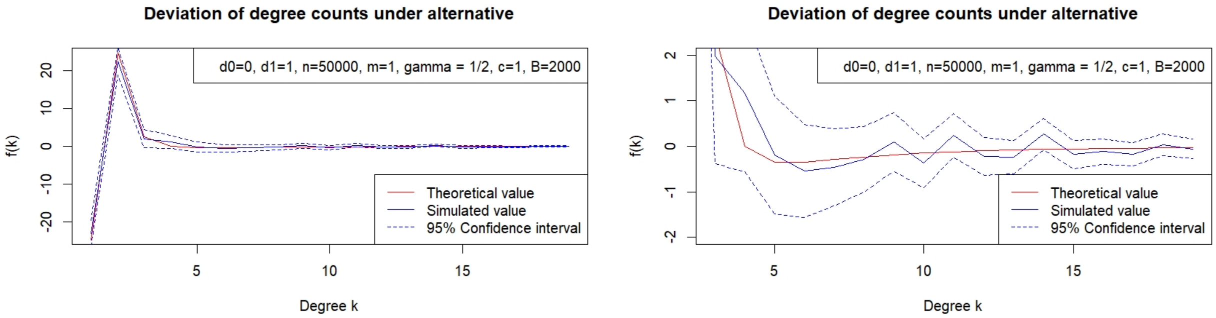


Figure 4: Deviation of the degree counts with $\delta_0 = 0, \delta_1 = 1, m = 1, n = 50000, c = 1, \gamma = \frac{1}{2}$.

In Figure 4 we see the simulations match the theoretical value better than in Figure 3. What we also see is that $f(k)$ is further away from zero for lower degrees. This indicates that the number of vertices with degree m, or degree $m+1$ is more useful than the number of vertices with a relatively high degree.

We now have an expression for the expected degree count under the alternative hypothesis for general k. We have seen that the proof of (7) consists of two main arguments. The first argument is that vertices introduced after the change point do not contribute significantly towards the final result. The second argument is that if a vertex has degree $k-2$ at the change point, the likelihood of that vertex having degree k in G_n is very small. In Section 3.2 we continue with the (co)variance of the degree counts.

3.2 Covariance of degree counts

With the expected number of vertices for general degree k under the alternative hypothesis we need one more tool to properly formulate the statistical test. The number of vertices with fixed degree k is a random variable. This means that $N_{k_1}(n)$ and $N_{k_2}(n)$ are in general correlated for all $k_1, k_2 \in \mathbb{N}$. If we want to formulate a test based on multiple degree counts, we need to take the correlation into account. The authors of [2] looked into the covariance matrix for general degree k. They found a very useful result that we state here. Before we state the result, we need to introduce some notation. For a random vector $(X_1^{(n)}, X_2^{(n)}, \dots)$, we write $(X_1^{(n)}, X_2^{(n)}, \dots) \overset{D}{\Rightarrow} (X_1, X_2, \dots)$ to show that for any $k \in \mathbb{N}$, as $n \to \infty$, $(X_1^{(n)}, X_2^{(n)}, \dots, X_k^{(n)})$ converges to $(X_1, X_2, \dots, X_k)$ in distribution. Now we can state the result from [2]. As $n \to \infty$,

$$\Big(\frac{N_k(n) - np_k(\delta_0)}{\sqrt{n}}, k = m, m+1, \dots\Big) \overset{D}{\Rightarrow} (Z_k, k = m, m+1, \dots),$$

where Z_k is a mean-zero Gaussian process with covariance function $R_Z(l, r)$. Furthermore, in Lemma 9.2 of [6] it is proven that even under the alternative hypothesis H_1 where $\delta(t)$ is a step function that

$$\Big(\frac{N_k(n) - \mathbb{E}_1[N_k(n)]}{\sqrt{n}}, k = m, m+1, \dots\Big) \overset{D}{\Rightarrow} (Z_k, k = m, m+1, \dots), \tag{14}$$

where Z_k is again a mean-zero Gaussian process with the same covariance function $R_Z(l, r)$. This shows that under the alternative model there is only a mean shift. The degree count retains the same distribution with the same covariance function. We omit the exact formula of R_Z in this thesis as it is very complicated, see [2] for the complete formula. By using their formula, we can formulate the asymptotic covariance matrix for the degree count under both hypotheses.

3.2.1 Two by two covariance matrix

The main focus is adding one new piece of information, the number of vertices with degree $m+1$. This means we only need to consider the two by two covariance matrix for $(N_m(n), N_{m+1}(n))$. This corresponds to three values for $R_Z(l,r)$; $l,r=m$; $l=m+1, r=m$ and $l=m+1, r=m+1$. We specialize the covariance formula in [2] to the relevant choices of l and m. For simplicity, let $\frac{\delta_0}{m}=w$. Then we have the following values in the covariance matrix:

$$
\begin{aligned}
&R_Z(m,m)\\
&=\frac{1}{(2m(m+\delta_0+1)+\delta_0)}\Big((m-1)(m+\delta_0)p_m(\delta_0)+\sum_{q=m+1}^{\infty}(q+\delta_0)p_q(\delta_0)\Big)\\
&-\frac{(2m+\delta_0)(m+1+\delta_0)^2}{(m+2+\delta_0+w)^2(2m(m+\delta_0+1)+\delta_0)}-\frac{(2m+\delta_0)(m-1)(\delta_0+m)^2}{m^2(m+2+\delta_0+w)^2(2m(m+\delta_0+1)+\delta_0)},
\end{aligned}
$$

$$
\begin{aligned}
&R_Z(m,m+1)\\
&=\Bigg(\frac{(m+\delta_0)^2(m-1)p_m(\delta_0)}{m(1+2(m+\delta_0+1)+w)(2(m+\delta_0+1)+w)}\\
&-\frac{(m+\delta_0)(m-1)p_m(\delta_0)}{m(1+2(m+\delta_0+1)+w)}\\
&+\frac{(m+1+\delta_0)(m+\delta_0)p_{m+1}(\delta_0)}{m(1+2(m+\delta_0+1)+w)(2(m+\delta_0+1)+w)}\\
&-\frac{(m+1+\delta_0)p_{m+1}(\delta_0)}{m(1+2(m+\delta_0+1)+w)}\\
&+\sum_{q=m+2}^{\infty}\frac{(q+\delta_0)(m+\delta_0)p_q(\delta_0)}{m(1+2(m+\delta_0+1)+w)(2(m+\delta_0+1)+w)}\Bigg)\\
&-(2m+\delta_0)(m+\delta_0)\Big(\frac{(m+1+\delta_0)^2}{(m+2+\delta_0+w)^2(2m(m+\delta_0+1)+\delta_0)}\\
&-\frac{(m+3+\delta_0-\frac{m+1}{m})(m+1+\delta_0)}{(m+3+\delta_0+w)(m+2+\delta_0+w)(2m(m+\delta_0+1)+\delta_0+m)}\Big)\\
&-\frac{(2m+\delta_0)(m-1)(m+\delta_0)}{m^2}\Big(\frac{(\delta_0+m)^2}{(m+2+\delta_0+w)^2(2m(m+\delta_0+1)+\delta_0)}\\
&-\frac{(\delta_0+m)(\delta_0+1+m)}{(m+3+\delta_0+w)(m+2+\delta_0+w)(2m(m+\delta_0+1)+\delta_0+m)}\Big),
\end{aligned}
$$

$$
\begin{aligned}
&R_Z(m+1,m+1)\\
&=\frac{(m+\delta_0)p_m(\delta_0)}{(2+2(m+\delta_0+1)+w)}\Big(\frac{2(m+\delta_0)^2(m-1)}{m(1+2(m+\delta_0+1)+w)(2(m+\delta_0+1)+w)}\\
&-\frac{2(m+\delta_0)(m-1)}{(1+2(m+\delta_0+1)+w))}+1\Big)\\
&+\frac{(m+\delta_0+1)p_{m+1}(\delta_0)}{(2+2(m+\delta_0+1)+w)}\Big(\frac{2(m+\delta_0)^2}{m(1+2(m+\delta_0+1)+w)(2(m+\delta_0+1)+w)}\\
&+\frac{2(m+\delta_0)}{m(1+2(m+\delta_0+1)+w)}+1\Big)\\
&+\sum_{q=m+2}^{\infty}\frac{2(m+\delta_0)^2(q+\delta_0)p_q(\delta_0)}{m}\frac{\Gamma(2(m+\delta_0+1)+w)}{\Gamma(3+2(m+\delta_0+1)+w)}
\end{aligned}
$$

$$
\begin{aligned}
&-(2m+\delta_0)(m+\delta_0)^2\Big(\frac{(m+1+\delta_0)^2}{(2m(m+\delta_0+1)+\delta_0)(m+2+\delta_0+w)^2}\\
&-\frac{2(m+1+\delta_0)(m+3+\delta_0-\frac{m+1}{m})}{(m+2+\delta_0+w)(m+3+\delta_0+w)(2m(m+\delta_0+1)+\delta_0+m)}\\
&+\frac{(m+3+\delta_0-\frac{m+1}{m})^2}{(m+3+\delta_0+w)^2(2m(m+\delta_0+1)+\delta_0+2m)}\Big)\\
&-\frac{(2m+\delta_0)(m-1)(m+\delta_0)^2}{m^2}\Big(\frac{(\delta_0+m)^2}{(m+2+\delta_0+w)(2m(m+\delta_0+1)+\delta_0)}\\
&-\frac{2(\delta_0+m)(\delta_0+m+1)}{(m+3+\delta_0+w)(m+2+\delta_0+w)(2m(m+\delta_0+1)+\delta_0+m)}\\
&+\frac{(\delta_0+m+1)^2}{(m+3+\delta_0+w)^2(2m(m+\delta_0+1)+\delta_0+2m)}\Big).
\end{aligned}
$$

In order to simplify these terms in the covariance matrix, we use a few facts. The biggest problems are the infinite sums in the expressions above. These can be simplified by noting the following two facts. The first fact is that $p_k(\delta_0)$ is the aymptotic fraction of vertices that have degree k. Therefore, if we sum over all possible $k \geq m$, we get the value 1. The second fact is the sum of all degrees present in the graph. If we sum the degree of every vertex this value will always be determined in advance. We start the model with two vertices connected with m edges, and we know that every vertex that gets introduced brings m edges. One edge connects two vertices, and we have $n+1$ vertices in total. Therefore the total degree of the graph after $n+1$ vertices will be $2mn$. Then the average degree of a single vertex is equal to the total degree divided by $n+1$, which is equal to $2m(\frac{n}{n+1})$. We again note that $p_k(\delta_0)$ is the fraction of vertices with degree k for $n \to \infty$. If we now sum $(kp_k(\delta_0))$ over all possible $k \geq m$, we have the expected degree of a single vertex, which is equal to $2m(\frac{n}{n+1})$. Of course we need to take the limit $n \to \infty$ for the asymptotic value as $p_k(\delta_0)$ is asymptotic. The limit gives us

$$\lim_{n\to\infty} 2m\left(\frac{n}{n+1}\right) = 2m.$$

These facts give us the tools to compute the infinite sums:

$$\sum_{q=m}^{\infty} p_q(\delta_0) = 1,$$
$$\sum_{q=m}^{\infty} q p_q(\delta_0) = 2m.$$

Using these facts, we then get

$$
\begin{aligned}
R_Z(m,m) &= \frac{m^2(m+\delta_0)(m+\delta_0+1)(2m+\delta_0)}{(2m(m+\delta_0+1)+\delta_0)(m^2+m\delta_0+2m+\delta_0)^2},\\
R_Z(m,m+1) &= (-1)\cdot\frac{m^2(m+\delta_0)(m+\delta_0+1)(2m+\delta_0)(2m(m+\delta_0+1)+\delta_0)^{-1}a(m,\delta_0)}{(2m(m+\delta_0+1)+\delta_0+m)(m^2+m\delta_0+2m+\delta_0)^2(m^2+m\delta_0+3m+\delta_0)},\\
a(m,\delta_0) &= m^4+7m^3+8m^2+2m^3\delta_0+11m^2\delta_0+m^2\delta_0^2+4m\delta_0^2+8m\delta_0+2\delta_0^2,
\end{aligned}
$$

$$
\begin{aligned}
&R_Z(m+1,m+1)\\
&=\frac{m^2(m+\delta_0)(m+\delta_0+1)(2m+\delta_0)(2m(m+\delta_0+1)+\delta_0)^{-1}b(m,\delta_0)}{(2m(m+\delta_0+1)+\delta_0+m)(2m(m+\delta_0+1)+\delta_0+2m)(m^2+2m+m\delta_0+\delta_0)^2(m^2+3m+m\delta_0+\delta_0)^2},
\end{aligned}
$$

where the polynomial $b(m,\delta_0)$ is given by

$$b(m,\delta_0) = 6\delta_0^4 m^4 + 18\delta_0^4 m^3 + 26\delta_0^4 m^2 + 17\delta_0^4 m + 4\delta_0^4$$

$$+ 24\delta_0^3 m^5 + 102\delta_0^3 m^4 + 178\delta_0^3 m^3 + 138\delta_0^3 m^2 + 38\delta_0^3 m$$
$$+ 36\delta_0^2 m^6 + 198\delta_0^2 m^5 + 432\delta_0^2 m^4 + 407\delta_0^2 m^3 + 134\delta_0^2 m^2 + 24\delta_0 m^7$$
$$+ 162\delta_0 m^6 + 434\delta_0 m^5 + 510\delta_0 m^4 + 208\delta_0 m^3 + 6m^8 + 48m^7 + 154m^6 + 224m^5 + 120m^4.$$

3.2.2 Simulation of covariance

The expressions above reflect the asymptotic values of the covariance matrix, meaning that they only apply to cases where $n \to \infty$. However, the formulas are already very accurate for large finite n. In this part we compare the theoretical values to a simulated instance with finite n. A second reason why we check the formulas from [2] with simulation is that the original version of this paper (published in Februari 2021) contained an error. During this thesis we found the error and communicated this error with the original authors. This lead to an updated version published in April 2024. For a more detailed explanation see Appendix A.

We simulate $B = 2000$ instances of a Preferential Attachment Model with $n = 10000$ vertices. We take different combinations of m and δ_0. We again use the algorithm in Appendix B, but this time we specify that there is no change point. Then the simulation will give us 2000 instances of $(N_m(n), N_{m+1}(n))$. We subtract the theoretical values $np_m(\delta_0)$ and $np_{m+1}(\delta_0)$ of the respective degree count and divide by $\sqrt{n}$. Finally we compute the sample variance and covariance. We can then compare this with the theoretical values. The results are in Table 1.

	Theoretical covariance matrix	Simulated covariance matrix
$m = 1, \delta_0 = 0$	$\begin{pmatrix} \frac{1}{9} & -\frac{4}{45} \\ -\frac{4}{45} & \frac{23}{180} \end{pmatrix}$	$\begin{pmatrix} 0.10900951 & -0.08714316 \\ -0.08714316 & 0.12843439 \end{pmatrix}$
$m = 5, \delta_0 = 0$	$\begin{pmatrix} \frac{5}{49} & -\frac{85}{1274} \\ -\frac{85}{1274} & \frac{1325}{10192} \end{pmatrix}$	$\begin{pmatrix} 0.0997287 & -0.06186127 \\ -0.06186127 & 0.12019103 \end{pmatrix}$
$m = 1, \delta_0 = 5$	$\begin{pmatrix} \frac{294}{3211} & -\frac{1554}{16055} \\ -\frac{1554}{16055} & \frac{2532}{16055} \end{pmatrix}$	$\begin{pmatrix} 0.09126148 & -0.09870273 \\ -0.09870273 & 0.16262993 \end{pmatrix}$
$m = 5, \delta_0 = -4$	$\begin{pmatrix} \frac{75}{484} & -\frac{75}{847} \\ -\frac{75}{847} & \frac{95175}{704704} \end{pmatrix}$	$\begin{pmatrix} 0.15223743 & -0.08810596 \\ -0.08810596 & 0.13452922 \end{pmatrix}$

Table 1: Comparison of theoretical and simulated covariance matrix for different values of m and δ_0.

Using a calculator we can get an idea of how close the simulated values are to the theoretical values. We give one example here. We have that $\frac{2532}{16055} \approx 0.157707879$, which deviates roughly 0.005 from the simulated value. We invite you to grab a calculator and check the other values as well.

In this chapter we have computed the asymptotic expected number of vertices with degree k under the alternative hypothesis. We showed both a proof for the asymptotic case, as well as simulated results to back up the equation. Then we proved a central limit theorem for the empirical degree counts under both the null hypothesis and the alternative hypothesis. We also used [2] to derive the asymptotic covariance matrix for the degree counts. Using this formula we have calculated the two by two covariance matrix for $(N_m(n), N_{m+1}(n))$. We now have all the information needed to introduce the test statistic based on both degree counts.

4 Generalized maximum likelihood ratio test

In Chapter 3 we have proven two key results that we need to introduce our new test statistic. In this chapter we first introduce the theory of hypothesis testing. This is done in Section 4.1. In Section 4.2 we introduce the generalized maximum likelihood ratio test. In Section 4.3 we formulate a new test to detect a change point with the generalized maximum likelihood ratio test.

4.1 Statistical testing

In the world around us, a lot of processes are random. If you count cars passing by your house, the number of cars that pass by in a certain time frame is random. In order to better understand processes like these, researchers first gather data from the random process. After obtaining a data sample of the process, a statistical test is used to decide if a data sample supports a particular hypothesis. If you gather a data sample from a random process, this data can reflect a particular distribution with an unknown parameter. It is well-known that the number of cars passing by in a certain time frame can be approximated by a Poisson distribution.

We now formally introduce a general hypothesis test. Let $\mathbb{X}$ be a random process with underlying distribution X. Let $f_\theta(x)$ be the probability mass/density function of X. The parameter θ lies in parameter space Θ, but is unknown. To make an informed statement about θ we formulate two distinct hypotheses. We have the hypothesis to be tested, i.e. the null hypothesis (H_0). We also have the alternative hypothesis (H_1). Formally we have a general hypothesis testing problem defined as

$$H_0 : \theta = \theta_0 \in \Theta_0 \qquad H_1 : \theta = \theta_1 \notin \Theta_0.$$

The null hypothesis states that $\theta = \theta_0 \in \Theta_0$ for some $\Theta_0 \subset \Theta$. The alternative hypothesis states that $\theta = \theta_1 \notin \Theta_0$. In order to determine which hypothesis is more likely to be true, we first gather independent data samples from random process $\mathbb{X}$. Using this data sample we construct an appropriate test statistic, say y. The test statistic $y = y(\boldsymbol{x})$ is a function of the data sample $\boldsymbol{x}$. We then formulate a function $\varphi : \mathbb{X} \mapsto \{0, 1\}$. Based on a data sample obtained from $\mathbb{X}$, the function φ will return either 0 if the data sample supports the null hypothesis, or 1 if the data sample does not support the null hypothesis. Most functions have the form

$$\varphi(\boldsymbol{x}) := \mathbb{1}\{y \in C\},$$

where the set $C \subset \Theta$ is called the critical region. If the test statistic $y \in C$, the function φ returns the value 1 and we reject the null hypothesis. We fail to reject the null hypothesis if $y \notin C$. Based on this hypothesis test there are four distinct outcomes. These are shown in Figure 5.

The first possible outcome is that the null hypothesis is true ($\theta = \theta_0$), but because the test statistic $y \in C$, we reject the null hypothesis. This is known as the type I error and is defined as

$$\mathbb{P}_0(\varphi(\boldsymbol{x}) = 1).$$

The second possible outcome is that the null hypothesis is true and the test statistic correctly fails to reject the null hypothesis. This is called the confidence of the test, and is defined as

$$\mathbb{P}_0(\varphi(\boldsymbol{x}) = 0).$$

Most researchers require the confidence of a test to be at least 95%. In some fields, like medicine, a higher confidence is required. Note that the type I error and the confidence of a test always add up to 1.
The third possible outcome is that the null hypothesis is false ($\theta = \theta_1$) and $y \in C$. Therefore we correctly reject the null hypothesis. This is known as the power of the test and is defined as

$$\mathbb{P}_1(\varphi(\boldsymbol{x}) = 1).$$

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Figure 5: Four outcomes of a hypothesis test [3].

The fourth and last possible outcome is that the null hypothesis is false, but because $y \notin C$ we fail to reject the null hypothesis. This is known as the type II error of the test, and is defined as

$$\mathbb{P}_1(\varphi(\boldsymbol{x}) = 0).$$

Note that the type II error and the power of the test always add up to 1. In a general statistical test we would like to keep the type I and type II error as low as possible. There is a trade-off between the type I and II error. In general, the lower the type I error the higher the type II error and vice versa. It is always possible to achieve a type I and type II error of zero by either never rejecting the null hypothesis or always rejecting the null hypothesis. However, this is not useful in practice. If we recall the example of the power grid in Chapter 1, it is desirable to find a good balance. If we use a statistical test to determine a problem in the power grid, we want to accurately detect a problem when it occurs. If there is no problem present, we also do not want to take preventive measures. In Section 4.2 we introduce a specific statistical test, the likelihood ratio test.

4.2 Likelihood-ratio test

As a statistical test, the generalized likelihood-ratio test is widely used to formulate criteria for rejecting or accepting a certain hypothesis. The generalized likelihood-ratio test is a generalization of the likelihood ratio test prescribed by the Neyman-Pearson Lemma [23]. Using a sample $\boldsymbol{x} = (x_1, \dots, x_n)$ of the random process $\mathbb{X}$ we can use the generalized maximum likelihood test to reject or fail to reject the null hypothesis. Under the null hypothesis, we state $\theta \in \Theta_0 \subset \Theta$ for some parameter space Θ. Under the alternative hypothesis we state $\theta \notin \Theta_0$. The test statistic for the generalized likelihood-ratio test λ is defined as

$$\lambda = -2 \ln \left(\frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)} \right). \tag{15}$$

The function $\mathcal{L}(\theta)$ is defined as the likelihood function. The likelihood function has two definitions, one for a discrete random variable, one for a continuous random variable. Let x be a realization of the random variable X. If X is discrete, the likelihood function $\mathcal{L}(\theta)$ is defined as

$$\mathcal{L}(\theta|x) = p_\theta(x) = \mathbb{P}_\theta(X = x).$$

If X is a continuous random variable, $\mathcal{L}(\theta)$ is defined as

$$\mathcal{L}(\theta|x) = f_\theta(x).$$

We give a short intuitive explanation of the likelihood function. If the function $\mathbb{P}_\theta(X = x)$ is a function of x with parameter θ fixed, it is the probability mass function of the random variable X. If the function $\mathbb{P}_\theta(X = x)$ is a function of θ with parameter x fixed, it is the likelihood function of X. The same idea holds when X is continuous. An example of a likelihood function is the likelihood function of a normal distribution. Assume both the mean μ and variance σ^2 are unknown. Suppose we have n realizations of a normal random variable $(x_1, \dots, x_n)$, then the likelihood function equals

$$\mathcal{L}(\mu, \sigma^2|x) = f_{\mu,\sigma^2}(x) = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}\right). \tag{16}$$

To be able to compute the generalized maximum likelihood test statistic we require an implicit estimate of the unknown parameters. If we take (16) to formulate a test, we still do not know the values for μ and σ^2. This is why we make an estimation based on the data. This educated guess is called the maximum likelihood estimator, and is defined as

$$\hat{\theta} = \underset{\theta\in\Theta}{\arg\max} \quad \mathcal{L}(\theta; \boldsymbol{x}), \tag{17}$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$ is the data sample.

4.3 Likelihood-ratio test for joint degree distribution

Using the results from Chapter 3 we can formulate a statistical test to determine the existence of a change point using the joint distribution of the number of vertices with degree m and the number of vertices with degree $m + 1$. Suppose we observe graph G_n. From the graph G_n we obtain $(N_m(n), N_{m+1}(n))$. We then subtract the empirical degree distributions $np_m(\delta_0)$ and $np_{m+1}(\delta_0)$ from the degree counts. We then divide by $\sqrt{n}$ and get

$$Y := \frac{1}{\sqrt{n}} \begin{pmatrix} N_m(n) - np_m(\delta_0) \\ N_{m+1}(n) - np_{m+1}(\delta_0) \end{pmatrix}.$$

If there is no change point (H_0), then Y should behave asymptotically like

$$Y \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right),$$
$$\Sigma = \begin{pmatrix} R_Z(m, m) & R_Z(m, m+1) \\ R_Z(m, m+1) & R_Z(m+1, m+1) \end{pmatrix}.$$

However, using the results of [6] and (7) we know the behaviour of Y under the alternative model. We know due to (14) that Y is still multivariate normal with covariance matrix Σ under the alternative model. The only difference between the null model and alternative model is the mean shift of Y. To compute the expected value of Y under the alternative, we fill in $k = m + 1$ in (7) and divide by $\sqrt{n}$ to obtain

$$\mathbb{E}_1[Y] = cn^{\gamma-\frac{1}{2}} \frac{\delta_1 - \delta_0}{(2 + \frac{\delta_1}{m})(m + \delta_0 + 2 + \frac{\delta_0}{m})} \begin{pmatrix} \mu_1(m, \delta_0) \\ \mu_2(m, \delta_0) \end{pmatrix}, \tag{18}$$
$$\mu_1(m, \delta_0) = -1,$$
$$\mu_2(m, \delta_0) = \frac{2(2 + \frac{\delta_0}{m})}{m + \delta_0 + 3 + \frac{\delta_0}{m}}.$$

Using our results we have shown that Y is multivariate normal with covariance structure Σ under both models for $n \to \infty$. This allows us to formulate an approximate model to test for the existence of a change point. As the results are asymptotic, this approximate model works best for a large number of vertices n.

We formulate the test as follows. We assume $Y \sim \mathcal{N}(\xi \cdot \mu, \Sigma)$ is multivariate normal with $k = 2$, μ, Σ known and $\xi \in \mathbb{R}$ unknown. To check whether the graph has a change point we test

$$H_0 : \xi = 0, \qquad\qquad H_1 : \xi \neq 0. \tag{19}$$

Although this information is not used in the test itself, it implies that under the alternative model

$$\xi = cn^{\gamma-\frac{1}{2}} \frac{\delta_1 - \delta_0}{(2 + \frac{\delta_1}{m})(m + \delta_0 + 2 + \frac{\delta_0}{m})}. \tag{20}$$

Equation (20) is the magnitude of the mean shift. The vector $\mu = (\mu_1(m, \delta_0); \mu_2(m, \delta_0))$ captures the direction of the mean shift. Note that μ only depends on m and δ_0, which are (assumed) known variables. This is why the direction of the mean shift is always known, and the only unknown parameter in the test is the magnitude of the mean shift ξ.

To test this hypothesis, we first need to define a suitable test statistic based on Y. Then we need to determine criteria based on this statistic to determine when we reject the null hypothesis. When looking at both hypotheses in (19), it is not immediately clear what the best test statistic is. The null hypothesis is simple, but the alternative hypothesis is two-sided and composite. This means that the alternative hypothesis does not state that ξ is equal to a specific value, it only states that $\xi \neq 0$. We know that the generalized likelihood ratio test is suitable for composite hypotheses. We use the generalized likelihood ratio test statistic, with the maximum likelihood estimator for ξ. We first derive the maximum likelihood estimator for the unknown parameter ξ. Recall the definition in Equation (17). We know that Y is multivariate normal, so we use that to first calculate the likelihood function $\mathcal{L}(\xi)$ as

$$\mathcal{L}(\xi) = (2\pi)^{-1} \det(\Sigma^{-1/2}) \exp\Big(-\frac{1}{2}(Y - \xi\mu)^T \Sigma^{-1} (Y - \xi\mu)\Big).$$

Using the likelihood function we calculate the maximum likelihood estimator for ξ. Notice that a regular exponential function ($f(x) = e^x$) is monotone, so we only need to maximize the exponent:

$$\begin{aligned} MLE(\xi) := \hat{\xi} &= \arg\max_{\xi\in\mathbb{R}} \left\{ -\frac{1}{2}(Y - \xi\mu)^T \Sigma^{-1} (Y - \xi\mu) \right\} \\ &= \arg\min_{\xi\in\mathbb{R}} \left\{ Y^T\Sigma^{-1}Y - 2\xi\mu^T\Sigma^{-1}Y + \xi^2\mu^T\Sigma^{-1}\mu \right\}. \end{aligned}$$

We first minimize by taking the derivative with respect to ξ. This yields

$$\frac{d}{d\xi}\Big(Y^T\Sigma^{-1}Y - 2\xi\mu^T\Sigma^{-1}Y + \xi^2\mu^T\Sigma^{-1}\mu \Big) = -2\mu^T\Sigma^{-1}Y + 2\xi\mu^T\Sigma^{-1}\mu.$$

Now we set the derivative equal to zero to obtain $\hat{\xi}$, i.e.,

$$\begin{aligned} -2\mu^T\Sigma^{-1}Y + 2\xi\mu^T\Sigma^{-1}\mu &= 0, \\ \hat{\xi} &= \frac{\mu^T\Sigma^{-1}Y}{\mu^T\Sigma^{-1}\mu}. \end{aligned} \tag{21}$$

We also have to take the derivative again to check if we found a local minimum. If the second derivative is positive, it means that the value we found is a local minimum. If the second derivative is negative, this indicates a local maximum. To check for the local minimum we use the following two facts:

- Matrix Σ is a covariance matrix, therefore it is semi-positive definite. This means that for any vector $\mu \in \mathbb{R}^2$: $\mu^T\Sigma\mu \geq 0$.

- If Σ is semi-positive definite, then Σ^{-1} is also semi-positive definite.

Using these facts we can check that we indeed found a local minimum:

$$\frac{d}{d\xi}\Big(-2\mu^T\Sigma^{-1}Y + 2\xi\mu^T\Sigma^{-1}\mu\Big) = 2\mu^T\Sigma^{-1}\mu \geq 0.$$

Now we have the maximum likelihood estimator of ξ based on the data. Then the generalized likelihood ratio test is formulated as (15). Filling this in yields

$$\begin{aligned}
\lambda &= -2\log\frac{\sup_{\xi=0}\exp\Big(-\frac{1}{2}(Y-\xi\mu)^T\Sigma^{-1}(Y-\xi\mu)\Big)}{\sup_{\xi\in\mathbb{R}}\exp\Big(-\frac{1}{2}(Y-\xi\mu)^T\Sigma^{-1}(Y-\xi\mu)\Big)} \\
&= -2\log\frac{\exp\Big(-\frac{1}{2}Y^T\Sigma^{-1}Y\Big)}{\exp\Big(-\frac{1}{2}(Y-\hat{\xi}\mu)^T\Sigma^{-1}(Y-\hat{\xi}\mu)\Big)} \\
&= Y^T\Sigma^{-1}Y - (Y-\hat{\xi}\mu)^T\Sigma^{-1}(Y-\hat{\xi}\mu) \\
&= 2\hat{\xi}\mu^T\Sigma^{-1}Y - \hat{\xi}^2\mu^T\Sigma^{-1}\mu \\
&= 2\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} - \frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \\
&= \frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu}.
\end{aligned}$$

We have calculated the new test statistic based on both degree counts. We analyze how this test statistic behaves, so we can formulate appropriate criteria for rejecting the null hypothesis. We state the following proposition. Under the null hypothesis,

$$\frac{\mu^T\Sigma^{-1}Y}{\sqrt{\mu^T\Sigma^{-1}\mu}} \sim \mathcal{N}(0,1). \tag{22}$$

We first note that Y is multivariate normal with mean zero and covariance matrix Σ. This means that we can rewrite Y using [16] into a normal random vector AZ, where $Z = (Z_1, Z_2)$, Z_i are independent standard normal random variables and $AA^T = \Sigma$. This gives us

$$\begin{aligned}
\mu^T\Sigma^{-1}Y &= \mu^T\Sigma^{-1}AZ \\
&= \mu^T(AA^T)^{-1}AZ \\
&= \mu^T(A^T)^{-1}A^{-1}AZ \\
&= \mu^T(A^T)^{-1}Z.
\end{aligned}$$

This proves that $\mu^T\Sigma^{-1}Y$ can be rewritten into a linear combination of independent standard normal random variables. Therefore $\mu^T\Sigma^{-1}Y$ is also normally distributed. Now we prove that (22) has mean equal to zero and variance equal to one. As both components of Y have mean zero, any linear combination has mean zero as well,

$$\mathbb{E}_0\left[\frac{\mu^T\Sigma^{-1}Y}{\sqrt{\mu^T\Sigma^{-1}\mu}}\right] = 0.$$

Next we prove that the variance is equal to 1. To prove this, we use the following property. For $M \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^{n\times 1}$,

$$Var\left(\boldsymbol{b^T}M\right) = \boldsymbol{b^T}Var(M)\boldsymbol{b}. \tag{23}$$

Then the variance of (22) is

$$
\begin{aligned}
Var\left(\frac{\mu^T\Sigma^{-1}Y}{\sqrt{\mu^T\Sigma^{-1}\mu}}\right) &= \frac{\left(\mu^T\Sigma^{-1}Var(Y)(\mu^T\Sigma^{-1})^T\right)}{(\sqrt{\mu^T\Sigma^{-1}\mu})^2} \\
&= \frac{\mu^T\Sigma^{-1}\Sigma\Sigma^{-1}\mu}{\mu^T\Sigma^{-1}\mu} \\
&= \frac{\mu^T\Sigma^{-1}\mu}{\mu^T\Sigma^{-1}\mu} = 1.
\end{aligned}
$$

This proves that $Z = \frac{\mu^T\Sigma^{-1}Y}{\sqrt{\mu^T\Sigma^{-1}\mu}}$ has a standard normal distribution under the null hypothesis. If we square Z we get the likelihood ratio test statistic back, therefore this statistic follows a chi-squared distribution with one degree of freedom. This fact allows us to always control the type I error. To set the type I error equal to (significance level) α we reject H_0 if

$$
\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \geq \chi^2_{\alpha;1}.
$$

A simple calculation shows us that with this criteria the type I error is indeed equal to α:

$$
\begin{aligned}
\mathbb{P}_0\left(\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \geq \chi^2_{\alpha;1}\right) &= \mathbb{P}_0\left(\frac{\mu^T\Sigma^{-1}Y}{\sqrt{\mu^T\Sigma^{-1}\mu}} \geq z_{\alpha/2}\right) + \mathbb{P}_0\left(\frac{\mu^T\Sigma^{-1}Y}{\sqrt{\mu^T\Sigma^{-1}\mu}} \leq -z_{\alpha/2}\right) \\
&= \mathbb{P}_0\left(Z \geq z_{\alpha/2}\right) + \mathbb{P}_0\left(Z \leq -z_{\alpha/2}\right) \\
&= \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}
$$

We now have the new test statistic, which uses more information than the asymptotically calibrated degree test as introduced in [6]. In Chapter 5 we will compare the performance of both tests to check whether the added information results in a better performance.

5 Comparison between hypothesis tests

The authors of [6] have done an extensive analysis of their minimal degree test. In this chapter we do the same with the joint distribution test, and then compare the results. We first derive the exact type II error of our joint distribution test in Section 5.1.1 In Section 5.1.2 we also derive the type II error of the calibrated minimal degree test, as the authors of [6] have not given an explicit expression. We compare them theoretically in Section 5.2, and after the theoretical analysis we simulate Preferential Attachment Models with a change point to compare both tests in Section 5.3. We not only use simulation to compare both tests, we also compare the simulated results to the theoretical results.

5.1 Type II error

We calculate the asymptotic type II error for both tests theoretically. We start with the new joint degree test in Section 5.1.1, then we also calculate the type II error for the calibrated test from [6] in Section 5.1.2. Before we compute the type II error for both tests, we need to note that the asymptotic type II error can only be computed when $\gamma = \frac{1}{2}$. Recall that the expected value of Y under the alternative hypothesis is still normally distributed, but shifts in the mean. Equation (18) showed the direction of the mean shift, as well as the magnitude of the mean shift. The direction is a vector that only depends on m and δ_0, but the magnitude of the mean shift (Equation (20)) depends on n. The type II error will also be dependent on the magnitude of the shift. If we want to compute the asymptotic type II error we implicitly take $n \to \infty$. Therefore we can only compute an exact equation for the type II error when the dependence on n gets eliminated. Equation (20) showed that the magnitude of the shift $\xi = O(n^{\gamma - \frac{1}{2}})$. If we take $\gamma = \frac{1}{2}$, we eliminate the dependence on n, allowing us to compute an exact type II error. For the remainder of the derivation we assume $\gamma = \frac{1}{2}$.

5.1.1 Joint degree test

We compute the type II error for the joint degree test. We have the same test statistic, but now we compute the statistic under the alternative hypothesis. To make computations easier we calculate the power of the test first:

$$\mathbb{P}_1\left(\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \geq \chi^2_{\alpha;1}\right) = \mathbb{P}_1\left(\mu^T\Sigma^{-1}Y \geq z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu}\right) + \mathbb{P}_1\left(\mu^T\Sigma^{-1}Y \leq -z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu}\right). \quad (24)$$

To derive this probability, we again use the fact that Y is a normal random vector. As proven in [16], this means we can write Y as $Y = \xi\mu + AZ$. Note that the mean is equal to $\xi\mu$ here, as we are calculating Y under the alternative model. We still have $Z = (Z_1, Z_2)$ independent standard normal random variables and A such that $\Sigma = AA^T$. We use this to obtain

$$\begin{aligned}
&\mathbb{P}_1\left(\mu^T\Sigma^{-1}Y \geq z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu}\right)\\
&= \mathbb{P}_1\left(\mu^T\Sigma^{-1}(\xi\mu + AZ) \geq z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu}\right)\\
&= \mathbb{P}_1\left(\xi\mu^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}AZ \geq z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu}\right)\\
&= \mathbb{P}_1\left(\mu^T(A^T)^{-1}A^{-1}AZ \geq z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu} - \xi\mu^T\Sigma^{-1}\mu\right)\\
&= \mathbb{P}_1\left(\mu^T(A^T)^{-1}Z \geq z_{\alpha/2}\sqrt{\mu^T\Sigma^{-1}\mu} - \xi\mu^T\Sigma^{-1}\mu\right)
\end{aligned}$$

$$= \mathbb{P}_1\left(Z \geq z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right) \tag{25}$$
$$= 1 - \Phi(z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu}).$$

At line (25) we used Equation (23), combined with the fact that $Var(Z) = I$. By combining Equation (24) in the correct way, we obtain

$$\mathbb{P}_1\left(\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \geq \chi^2_{\alpha;1} \right) = \mathbb{P}_1\left(\mu^T\Sigma^{-1}Y \geq z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right) + \mathbb{P}_1\left(\mu^T\Sigma^{-1}Y \leq -z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right)$$
$$= 1 - \Phi\left(z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right) + \Phi\left(-z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right).$$

The expression above is the power of the test, so we adjust to receive the type II error.

$$\mathbb{P}_1\left(\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \leq \chi^2_{\alpha;1} \right) = 1 - \mathbb{P}_1\left(\frac{(\mu^T\Sigma^{-1}Y)^2}{\mu^T\Sigma^{-1}\mu} \geq \chi^2_{\alpha;1} \right)$$
$$= \Phi\left(z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right) - \Phi\left(-z_{\alpha/2} - \xi\sqrt{\mu^T\Sigma^{-1}\mu} \right). \tag{26}$$

5.1.2 Calibrated minimal degree test

To compare results, we consider the asymptotically calibrated minimal degree test from [6]. We have $T(G_n) = N_m(n) - np_m(\delta_0)$. Note that the variance used in [6] is equal to the variance derived in this thesis, $R_Z(m,m) = w(\delta_0, m)$. We first compute the power. We reject H_0 if

$$|T(G_n)| \geq \sqrt{nw(\delta_0, m)}z_{\alpha/2}.$$

Then under the alternative

$$\mathbb{P}_1\left(|T(G_n)| \geq \sqrt{nw(\delta_0, m)}z_{\alpha/2} \right) = \mathbb{P}_1\left(T(G_n) \geq \sqrt{nw(\delta_0, m)}z_{\alpha/2} \right) + \mathbb{P}_1\left(T(G_n) \leq -\sqrt{nw(\delta_0, m)}z_{\alpha/2} \right). \tag{27}$$

Note that $T(G_n) = \sqrt{n} \cdot Y_1$. We rewrite the expression above as

$$\mathbb{P}_1\left(T(G_n) \geq \sqrt{nw(\delta_0, m)}z_{\alpha/2} \right)$$
$$= \mathbb{P}_1\left(Y_1 \geq \sqrt{w(\delta_0, m)}z_{\alpha/2} \right)$$
$$= \mathbb{P}_1\left(\frac{Y_1 - \mathbb{E}[Y_1]}{\sqrt{R_Z(m,m)}} \geq z_{\alpha/2} - \frac{\mathbb{E}[Y_1]}{\sqrt{R_Z(m,m)}} \right)$$
$$= \mathbb{P}_1\left(Z \geq z_{\alpha/2} - \frac{(-\xi)}{\sqrt{R_Z(m,m)}} \right)$$
$$= 1 - \Phi\left(z_{\alpha/2} + \frac{\xi}{\sqrt{R_Z(m,m)}} \right).$$

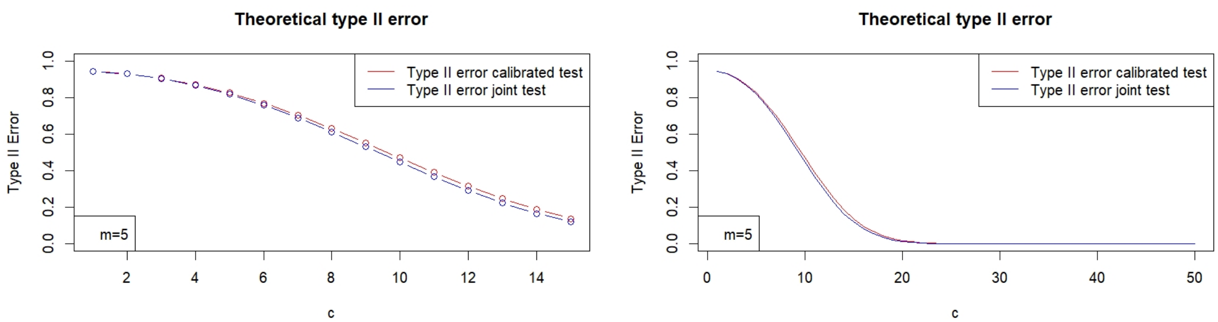
Then we compute the type II error by taking the complement of (27). This gives us

$$\mathbb{P}_1\left(|T(G_n)| \leq \sqrt{nw(\delta_0, m)}z_{\alpha/2} \right) = \Phi\left(z_{\alpha/2} + \frac{\xi}{\sqrt{R_Z(m,m)}} \right) - \Phi\left(-z_{\alpha/2} + \frac{\xi}{\sqrt{R_Z(m,m)}} \right). \tag{28}$$

We have computed the exact theoretical type II errors for both tests. In Section 5.2 we will compare them to each other theoretically, then in Section 5.3 we perform simulations to compare both tests, as well as compare the simulated results to the theoretical results.

5.2 Performance comparison

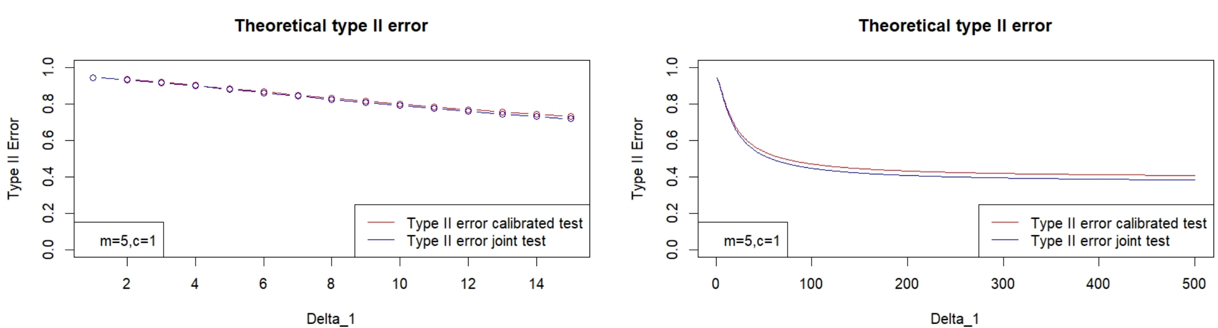
We compare Equations (26) and (28). In this section we will compare them theoretically, in Section 5.3 we will compare them using a simulation. We take $\alpha = 0.05$ for every comparison. Recall that $\gamma = \frac{1}{2}$. We first look at the theoretical comparison of both tests. We take $m = 5$ and $\delta_0 = 0$, and plot the theoretical type 2 error for different values of c and δ_1.



(a) $c \in [1, 15]$ (b) $c \in [1, 50]$

Figure 6: Type II error for various values of c, $m = 5$, $\delta_0 = 0$, $\delta_1 = 1$, $\gamma = \frac{1}{2}$ and $\alpha = 0.05$.

We see in Figure 6b that for c large enough the type II error will become zero. As c directly influences at what time the step function $\delta(t)$ changes value this is to be expected. We also see that the test based on the joint degree count $(N_m(n), N_{m+1}(n))$ performs slightly better than the asymptotically calibrated test. While it performs better, this difference is negligible. We also check the type II errors for other parameters. We first plot the type II error for various values of δ_1.



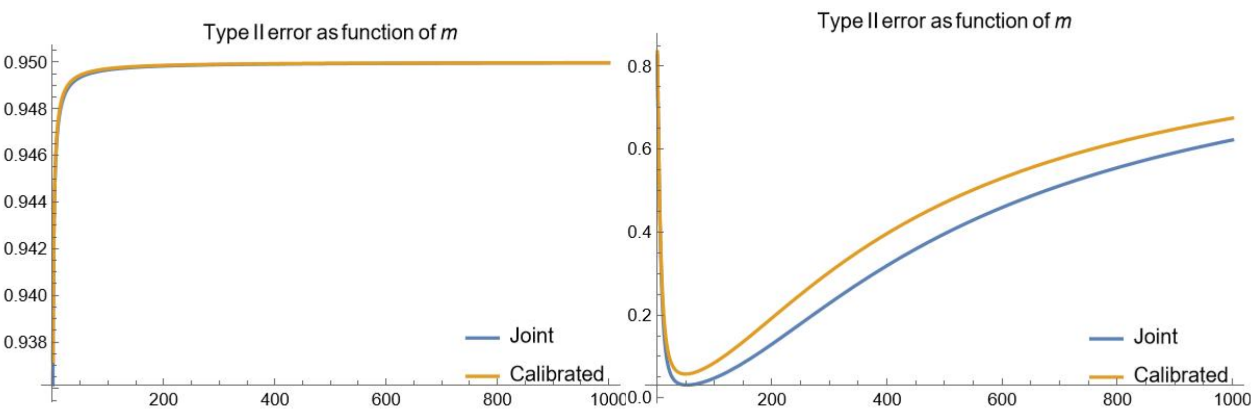
(a) $\delta_1 \in [1, 15]$ (b) $\delta_1 \in [1, 500]$

Figure 7: Type II error for various values of δ_1, $m = 5$, $\delta_0 = 0$, $c = 1$, $\gamma = \frac{1}{2}$ and $\alpha = 0.05$.

In Figure 7b we see more interesting behaviour of the type II error. The joint distribution test again performs ever so slightly better. However, when the other parameters are fixed, the magnitude of the change point does not result in a vanishing type II error. On the right side we see a plot where $\delta_1 \in \{0, \dots, 500\}$. Even for $\delta_1 = 500$, the type II error is still around 40 percent for both tests. When we analyze ξ, it becomes clear why. We know the implied value for ξ under H_1, recall Equation (18). Now let $\delta_1 \to \infty$ to obtain

$$\begin{aligned}\lim_{\delta_1\to\infty} \xi &= \lim_{\delta_1\to\infty} cn^{\gamma-\frac{1}{2}} \frac{\delta_1-\delta_0}{(2+\frac{\delta_1}{m})(m+\delta_0+2+\frac{\delta_0}{m})} \\ &= \lim_{\delta_1\to\infty} cn^{\gamma-\frac{1}{2}} \frac{1-\frac{\delta_0}{\delta_1}}{(\frac{2}{\delta_1}+\frac{1}{m})(m+\delta_0+2+\frac{\delta_0}{m})} \\ &= cn^{\gamma-\frac{1}{2}} \frac{m}{m+\delta_0+2+\frac{\delta_0}{m}}.\end{aligned}$$

For δ_1 approaching infinity, ξ will become constant. This means that the type II error will become constant as well. The last plot we show is the type II error as function of m.



(a) $\delta_1 = 1$ (b) $\delta_1 = 100$

Figure 8: Type II error for various values of m, $\delta_0 = 0$, $\delta_1 \in \{1, 100\}$, $c = 1$, $\gamma = \frac{1}{2}$ and $\alpha = 0.05$.

On the left-hand side we see that the type II error is increasing in m. If we take a different situation where $\delta_1 = 100$ in Figure 8b, we see that the type II error is not necessarily increasing in m. There is a minimum, and this minimum is attained at $m \approx \frac{\delta_1}{2}$. Figure 8b shows that the performance of both tests is heavily dependent on the model itself. Both tests show the same type of behaviour. This suggests that regardless of which test is chosen there is an incentive to choose the parameters of the model carefully. If you have influence over either m, δ_0 or δ_1, Figure 8b shows that choosing these parameters carefully can dramatically increase or decrease the power of both tests. Unfortunately in practice it is not always possible to influence these parameters.

5.3 Simulation results

To check the theoretical results in the previous section, we use the simulation in Appendix B to simulate the type I and type II errors for both tests. As the simulation only gives us data in the form of $(N_m(n), N_{m+1}(n))$, we can use one simulation to calculate the type I and type II error of both tests. To do this we generate $B = 2000$ instances of a Preferential Attachment Model without a change point, after that we generate $B = 2000$ instances of a Preferential Attachment Model with a change point. We then compute the test statistic for both the joint distribution test, as well as the calibrated test. We count the number of times both tests reject the null hypothesis when there is no change point or fails to reject the null hypothesis when there is a change point and divide this by B. This gives the simulated type I and type II error for both tests. We can then compare them. We start with the type I error in Section 5.3.1, in Section 5.3.2 we compare the type II error.

5.3.1 Type I error

We know that for significance level α both tests have an asymptotic type I error equal to α. Here we compare this theoretical result to a simulation. We start with $\alpha = 0.05$, $n = 10000$, $m = 5$, $\delta_0 = \delta_1 = 0$ and $\gamma = \frac{1}{2}$. We simulate for various values of c, but as there is no change point we essentially simulate 25 groups of 2000 independent Preferential Attachment Models.

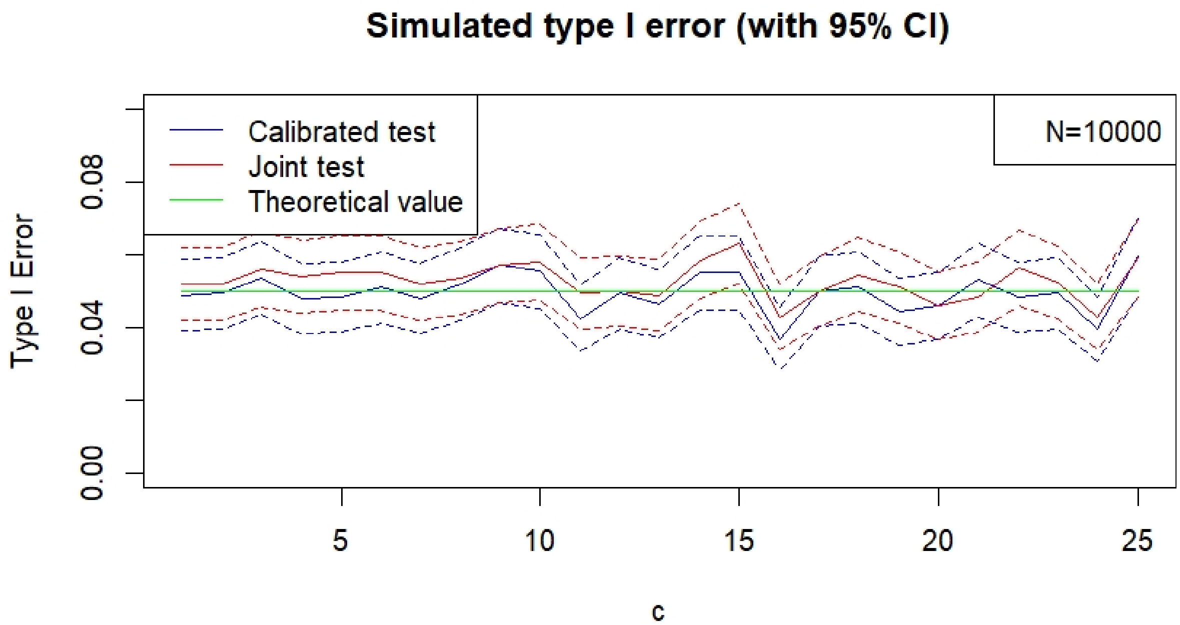


Figure 9: Simulated type I error for various values of c, $\alpha = 0.05$, $n = 10000$, $m = 5$, $\delta_0 = \delta_1 = 0$, $\gamma = \frac{1}{2}$.

Figure 9 is a bit cluttered, so we also plot the type I error of both tests separately in Figure 10.

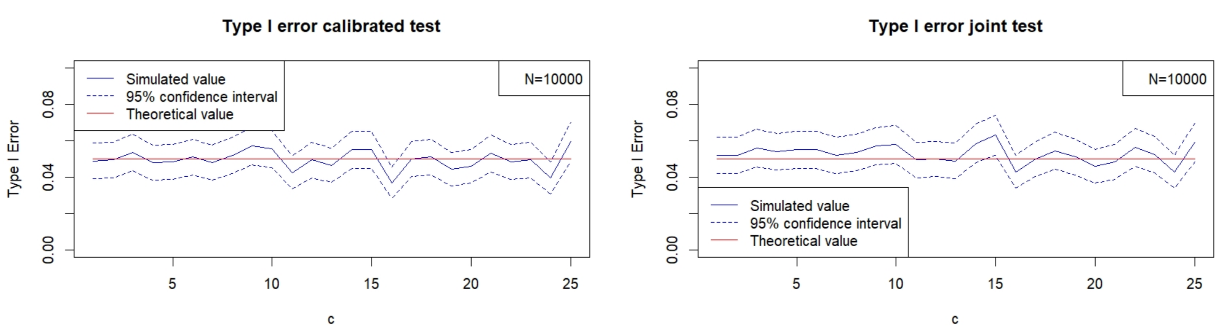


Figure 10: Theoretical and simulated type I error for various values of c, $\alpha = 0.05$, $n = 10000$, $m = 5$, $\delta_0 = \delta_1 = 0$, $\gamma = \frac{1}{2}$.

The simulation shows that the theoretical value and the simulation match very well. There are very few runs where the theoretical value lies outside the 95% confidence interval. We do notice that the simulated type I error is generally higher for the joint distribution test. Figure 9 shows that the joint distribution test rejects the null hypothesis more often than the calibrated minimal degree test. We do have to note that this difference is insignificant. We provide one more plot where we increase the number of vertices. We simulate $B = 2000$ runs for six different numbers of vertices: $n \in \{1000, 5000, 10000, 20000, 50000, 100000\}$.

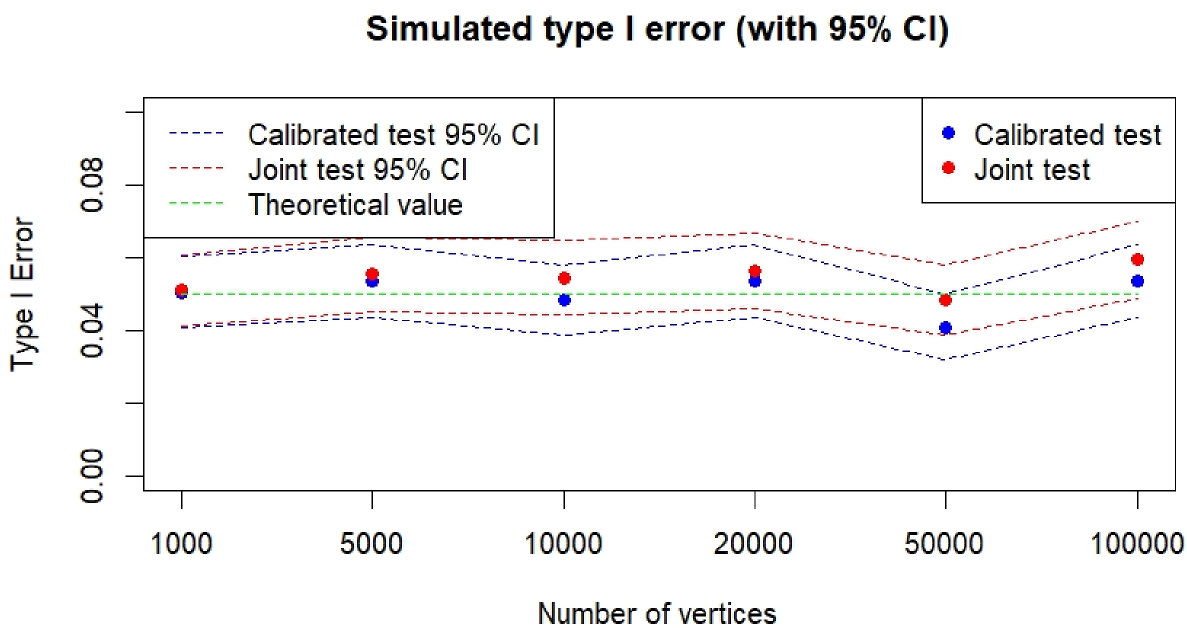


Figure 11: Simulated type I error for various values of n, $\alpha = 0.05$, $c = 1$, $m = 5$, $\delta_0 = \delta_1 = 0$, $\gamma = \frac{1}{2}$.

In Figure 11 we see that there is no notable effect when the number of vertices gets increased. We only see the same effect as in Figure 9, where the joint distribution test often has a larger type I error than the calibrated minimal degree test. The difference between both tests is negligible for all number of vertices. In the next section we look at simulations of the type II error.

5.3.2 Type II error

To simulate the type II error, we start with $n = 10000$ vertices, $m = 5$, $\delta_0 = 0$, $\delta_1 = 1$ and $\gamma = \frac{1}{2}$. We take $\alpha = 0.05$ for all simulations. We first plot the simulated type II error for both tests in Figure 12.

The simulation shows the same behaviour as the theoretical analysis. In Figure 12 both tests have similar type 2 errors, but the joint test that uses more information is always a few percent points better. In Figure 13 we plot the simulated type II error together with the theoretical value.

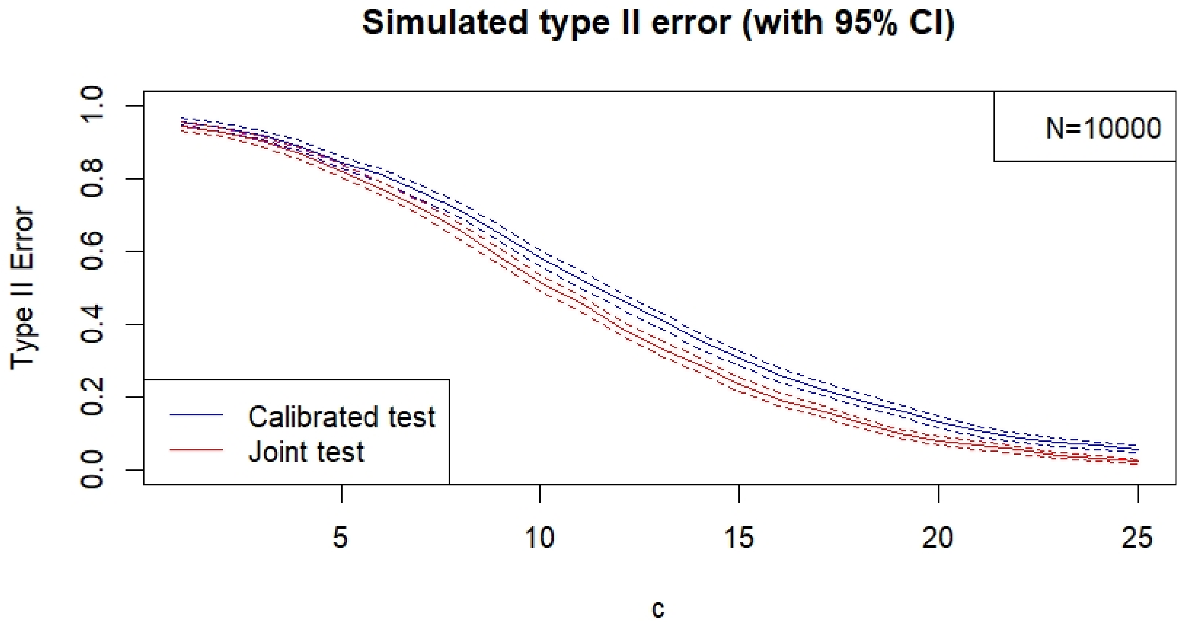


Figure 12: Simulated type II error for various values of c, $\alpha = 0.05$, $n = 10000$, $m = 5$, $\delta_0 = 0$, $\delta_1 = 1$, $\gamma = \frac{1}{2}$.

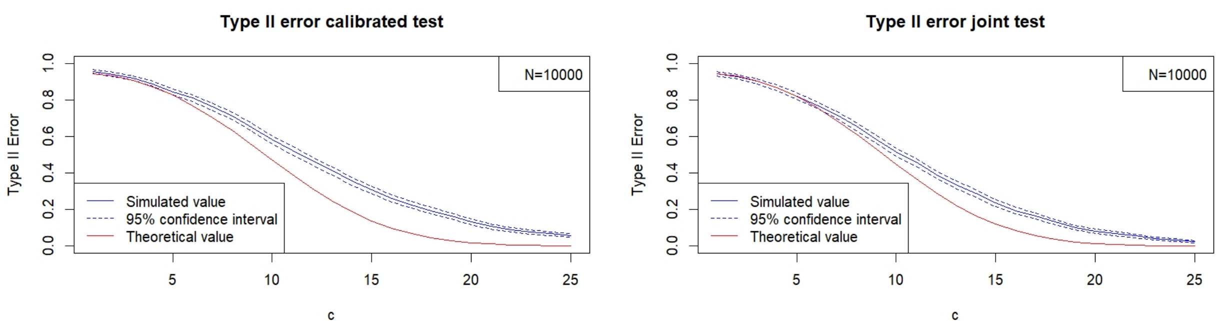


Figure 13: Theoretical and simulated type II error for various values of c, $\alpha = 0.05$, $n = 10000$, $m = 5$, $\delta_0 = 0$, $\delta_1 = 1$, $\gamma = \frac{1}{2}$.

When comparing the theoretical value and the simulation, there is a noticeable difference. This difference can be explained by the fact that the theoretical analysis leaves out a number of terms with order smaller than $O(n^\gamma)$. The simulation above is done using $n = 10000$, so the smaller order terms are still contributing slightly to the type II error. The final thing we check is the behaviour of the simulation compared to the theoretical value when the number of vertices n is increased. As $n \to \infty$, the simulated values should converge to the theoretical value. We plot this in Figure 14.

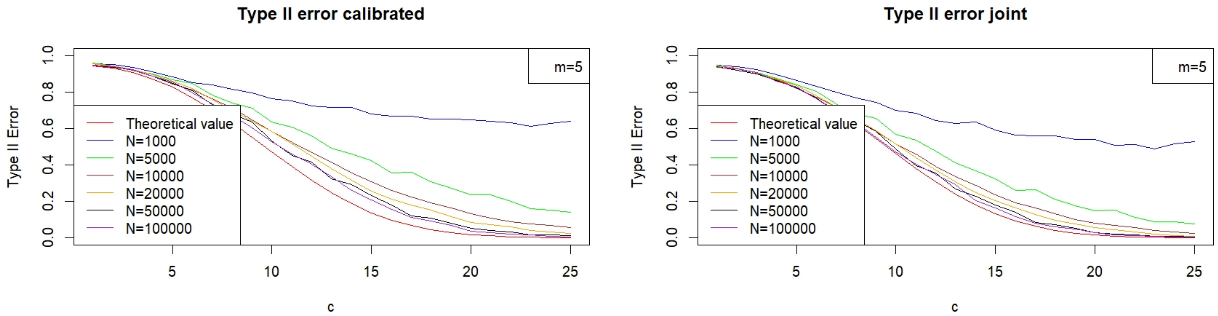


Figure 14: Theoretical and simulated type II error for various values of c and n, $\alpha = 0.05$, $m = 5$, $\delta_0 = 0$, $\delta_1 = 1$, $\gamma = \frac{1}{2}$.

When simulating for different values of the number of vertices n, we can clearly see behaviour towards the theoretical value. For $n = 1000$ the number of vertices is too small to be able to reliably detect a change point at all. As n increases we see the accuracy increase as well. When the theoretical type II error is large the difference is not very noticeable, but as we alter the location of the change point by increasing c, the simulations clearly show that the accuracy of the test increases when the number of vertices n goes up. The last thing to note is that the simulations show that the convergence speed towards the theoretical value for the joint distribution test is slightly faster than the convergence speed of the calibrated test. If we look back at Figure 13 where the number of vertices is equal, we see that the difference between the theoretical and simulated value is lower for the joint distribution test.

In this chapter we have first calculated the asymptotic equation for the type II error for both tests. Using these equations we have seen that the test based on the joint distribution $(N_m(n), N_{m+1}(n))$ is slightly better than the test based on only $N_m(n)$. The joint degree test performed better in both the theoretical analysis as well as the simulation. We do have to note that the differences between both tests are marginal. The type II error is a few percent points better for the joint degree test.

We have also seen that some parameters are more influential than others. As Figure 6b showed, the type II error will tend to zero when c is large enough. Figure 7b showed that δ_1 has less influence on the type II error. Finally when plotting for m we saw that the type II error heavily depends on the given parameters. For $\delta_0 = 0$, Figure 8b showed that the optimal choice of parameters is when $m \approx \frac{\delta_1}{2}$. In the simulation results we saw that the simulated type II error is always larger than the theoretical value, but that this value converges to the theoretical value as n increased. The simulation also showed that the type I error of the joint degree test is slightly higher than the type I error of the minimal degree test.

6 Estimating the location of the changepoint

One more thing that is interesting to investigate is the location of the change point. Using the test formulated in [6] as well as the updated test in this thesis we know that we can detect a change point reliably under the right circumstances. However, the test does not tell us anything about the change point. It only tells us that there exists one. In this final chapter we shortly introduce and analyze a way to estimate where the change point occurred in the creation process of the graph G_n.

6.1 Estimator for the change point

The location of the change point τ_n is parameterized as $\tau_n = n - cn^{\gamma}$. Under the assumption that you know δ_0 and δ_1, you can estimate the location of the change point using data from a graph. Note that knowledge of both δ_0 and δ_1 is generally rare. They are hard to estimate, so in practice this is only a viable way to estimate the location of the change point if you know both values exactly. We always know the value of n, so we will introduce an estimator for $\beta = cn^{\gamma}$. We know that the value cn^{γ} is captured in Equation (18). Recall that

$$\begin{aligned}
\mathbb{E}[Y] &= cn^{\gamma-\frac{1}{2}} \frac{\delta_1 - \delta_0}{(2+\frac{\delta_1}{m})(m+\delta_0+2+\frac{\delta_0}{m})} \begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \\
\mu_1 &= -1, \\
\mu_2 &= \frac{2(2+\frac{\delta_0}{m})}{m+\delta_0+3+\frac{\delta_0}{m}}.
\end{aligned}$$

Using this expectation, we can directly estimate the quantity $\beta = cn^{\gamma}$. Given that we know m, δ_0 and δ_1, we can formulate two estimators for β based on degree counts m and $m+1$.

$$\begin{aligned}
\hat{\beta}_1 &= \sqrt{n} Y_1 \frac{(2+\frac{\delta_1}{m})(m+\delta_0+2+\frac{\delta_0}{m})}{\delta_0 - \delta_1}, \\
\hat{\beta}_2 &= \sqrt{n} Y_2 \frac{(2+\frac{\delta_1}{m})(m+\delta_0+2+\frac{\delta_0}{m})(m+\delta_0+3+\frac{\delta_0}{m})}{2(\delta_1-\delta_0)(2+\frac{\delta_0}{m})}.
\end{aligned}$$

These estimators are based solely on the minimal degree count, or on the degree count of $m+1$. They do not use the combined information available. Therefore we can introduce one more estimator that tries to capture all the information available. In the formulation of the test based on joint degree counts we used the maximum likelihood estimator for ξ, recall Equation (21). We can use this estimator to estimate the exact value of $\beta = cn^{\gamma}$ as

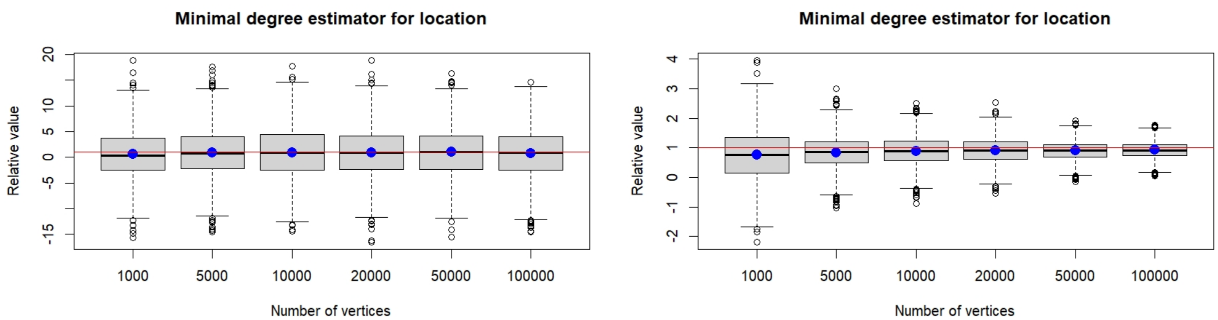
$$\begin{aligned}
\hat{\beta}_3 &= \hat{\xi} \frac{\sqrt{n}(2+\frac{\delta_1}{m})(m+\delta_0+2+\frac{\delta_0}{m})}{\delta_1 - \delta_0} \\
&= \frac{\mu^T R^{-1} Y}{\mu^T R^{-1} \mu} \cdot \frac{\sqrt{n}(2+\frac{\delta_1}{m})(m+\delta_0+2+\frac{\delta_0}{m})}{\delta_1 - \delta_0}.
\end{aligned}$$

6.2 Analysis of the estimators

We use simulation to analyze the performance of all three estimators. We simulate six instances of a Preferential Attachment Model with varying number of vertices, again using the simulation in Appendix (B). We take $n \in \{1000, 5000, 10000, 20000, 50000, 100000\}$. Then for every n we simulate $B = 2000$ runs. Because the value β is dependent on n, we can not look at the difference between the estimator and the true value. For large n this difference will always be larger than for small n. Therefore we calculate the following value for every estimator. For $i \in \{1, 2, 3\}$,

$$\text{Relative performance of } \hat{\beta}_i = \frac{\hat{\beta}_i}{cn^\gamma}.$$

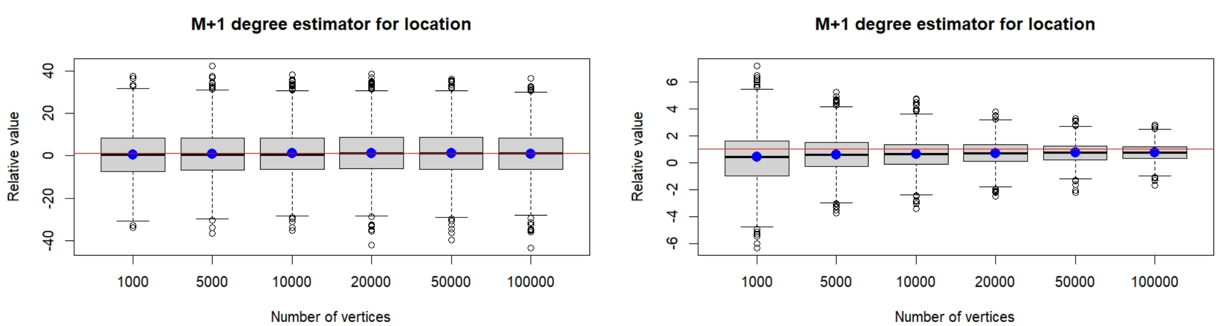
This means that for an estimator to be considered 'good', it needs to be as close to the value 1 as possible. Below are plots for the values of the relative performance. The red line in each plot indicates the value 1, the blue dot in each plot indicates the mean of every group.



(a) $\gamma = \frac{1}{2}$ (b) $\gamma = \frac{3}{4}$

Figure 15: Relative performance of $\hat{\beta}_1$ for various n, $\delta_0 = 0$, $\delta_1 = 1$, $c = 1$

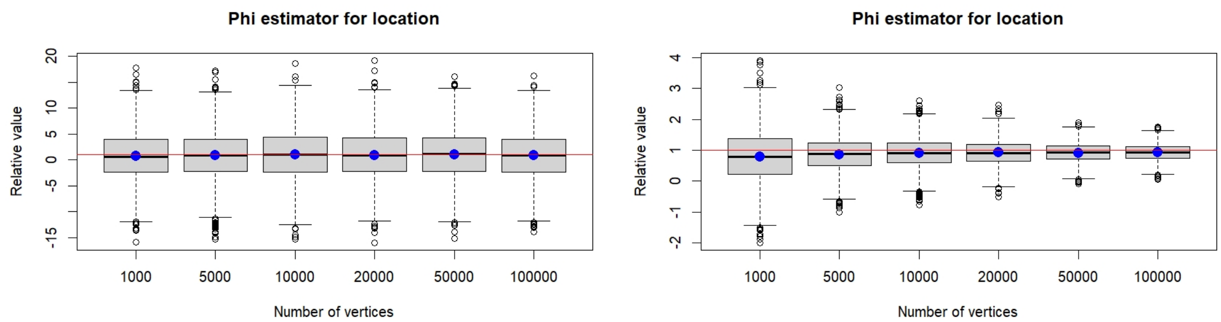
In Figure 15a we see that the estimator performs bad for $\gamma = \frac{1}{2}$. While the average value is close to 1, there are a lot of values exceeding 10. In Figure 15b we see that when $\gamma = \frac{3}{4}$ the performance of the estimator is better for every value of n. Here we also see that the performance increases when n is increased.



(a) $\gamma = \frac{1}{2}$ (b) $\gamma = \frac{3}{4}$

Figure 16: Relative performance of $\hat{\beta}_2$ for various n, $\delta_0 = 0$, $\delta_1 = 1$, $c = 1$

When analyzing the second estimator, we see that its performance is a lot worse than the first estimator. We also see the same behaviour as for the first estimator. For $\gamma = \frac{1}{2}$ the estimator does not get better with increasing n, for $\gamma = \frac{3}{4}$ it does get better.



(a) $\gamma = \frac{1}{2}$ (b) $\gamma = \frac{3}{4}$

Figure 17: Relative performance of $\hat{\beta}_3$ for various n, $\delta_0 = 0$, $\delta_1 = 1$, $c = 1$

The third estimator is based on the maximum likelihood estimator for ξ. In Figure 17a we see that the performance of $\hat{\beta}_3$ is very similar to $\hat{\beta}_1$. The performance also does not improve as n increases, but this is to be expected when $\gamma = \frac{1}{2}$. When $\gamma = \frac{3}{4}$, we see that the estimator becomes better for larger n, but there is still no visible difference between $\hat{\beta}_1$ and $\hat{\beta}_3$.

7 Conclusion and discussion

In Chapter 7 we shortly discuss the key findings in this thesis, as well as possibilities for future research.

7.1 Findings in the thesis

In this thesis we expanded a few key results from [6]. In Chapter 2 we looked at how a Preferential Attachment Model is defined. We looked at the relation between $p_k(\delta_0)$, the asymptotic fraction of vertices with degree k and the number of vertices with degree k in particular. The authors of [18] showed that the fraction of vertices with degree k is always concentrated around $p_k(\delta_0)$, recall (2). This is an important result that is extensively used in the test introduced in [6] as well as the test in this thesis.

In Chapter 3 we first analyzed how the degree counts change under the alternative hypothesis. By using the available research we were able to extend a key result from [6]. We were able to characterize the expected value of the number of vertices with degree k for general fixed k. The authors of [6] only analyzed this expression for $k = m$, where m is the minimal degree of any vertex in the graph. Recall the definition of the change point in Equation (3). Under the assumption that $\gamma < 1$, we now know the behaviour for any fixed $k \in \mathbb{N}$ with the constraint that $k = o(n^{1-\gamma})$. If k becomes too large, it will become a vertex that receives 'preferential attachment', which means that every new vertex introduced has a big chance of connecting to it. In this case our results do not hold. In the process we discovered two interesting facts:

1. The first fact is that the contribution of the vertices arriving after the change point τ_n do not contribute significantly towards the expected number of vertices with certain degree $k > m$ under the alternative model. Recall that for $n \to \infty$ these contributions could be ignored.
2. The second fact is that the probability that a vertex with degree equal to $k - 2$ or less at the change point has a very low chance of obtaining degree k in the final version of the graph, recall Proposition 2 (12). This can be seen as symptom of the 'rich get richer' phenomenon present in Preferential Attachment Models. Vertices with low degree have a low chance of receiving a higher degree after the change point.

After that we looked into the correlation between degree counts, already studied in [2]. We were able to derive exact formulas for the (co)variance of the degree counts. We also found a mistake in the original formula of [2], leading to an updated version by the original authors.

In Chapter 4 we were able to formulate the same hypothesis test, but with a test statistic based on the joint degree count of m and $m + 1$ using these results. Our results in Chapter 5 showed that adding this extra piece of info improves the performance of the test by a few percent points. We were able to compare this after calculating the asymptotic theoretical type II error for both tests, recall Equations (26) and (28). In [6] the authors proved that the type II error always vanishes when $\frac{1}{2} < \gamma < 1$ In this thesis we were able to compute an exact equation for the type II error under the assumption that $\gamma = \frac{1}{2}$. We also saw that the test performs very differently for different parameter settings. Different parameters had different effects on the type II error. The most interesting parameter was m itself. Figure 8b showed that m can have a drastic influence on the type II error. One more important thing to note is that the simulation results in Chapter 5 showed a tiny difference in the type I error. Both tests have the same type I error in theory, but the simulation showed that the calibrated minimal degree test is ever so slightly more conservative.

When we translate these results to real-life problems, this is not an issue. We recall the example given in the introduction of the thesis, the blackout in the United States on August 14 2003. If we assume that a statistical test is used to find anomalies in the power grid, a low false negative is preferable over a low false positive. The consequences of shutting down a small part of the power grid to save the other systems in the network are smaller than the consequences of a complete failure of the power grid. Other areas where statistical tests are used also prefer lower false negatives. As [20] states, false negatives pose greater risks to society that false positives.

Finally in Chapter 6 we looked at estimators for the exact location of the change point. We introduced three estimators. Two estimators are based on the degree counts of m and $m+1$, and one estimator is based on the maximum likelihood estimator derived for the test in Chapter 4. Using the simulation, we could see that under the regime of $\gamma = \frac{3}{4}$ the estimators become more accurate when n is increased. When taking $\gamma = \frac{1}{2}$, the estimators have a huge variance when predicting the value $\beta = cn^{\gamma}$.

7.2 Future research

In future research, this thesis provides multiple leads. The first topic for further research is another update to the test statistic used in the hypothesis test. Adding one new piece of information improved the performance of the test. We would theorize that adding another piece of information will again improve the performance.

The second topic of future research is a key result shown in Figure 8b. In this figure it was shown that the behaviour of both test statistics are heavily dependent on the model choice. This might be an indication that the test performs optimal under very specific conditions. Future research can analyze these exact conditions. This can lead to a better understanding of the Preferential Attachment Model with a change in parameter. Because this thesis has already calculated the theoretical type II error for both the calibrated test and the minimal degree test, we only have to analyze these equations for different values of $n, m, \delta_0, \delta_1, c$ and γ. This can also lead to a proof of Conjecture 3.2 in [6]. In Conjecture 3.2, the authors theorized that any test based on degree counts becomes powerless when $\gamma < \frac{1}{2}$. Using the formulas provided in this thesis you can calculate the approximate type II error for instances where $\gamma < \frac{1}{2}$ and prove or disprove Conjecture 3.2.

More topics are also possible. In this thesis the function $\delta(t)$ is assumed to be a step function. In future research we can assume that $\delta(t)$ is a more complicated function. The last topic we discuss is the analysis of estimators for the change point τ_n. If we could devote more time in the thesis we would be able to analyze these estimators more closely. In this thesis we briefly looked at three estimators using the simulation. The first thing that we could do is analyze the estimators theoretically. Are the estimators biased? Are the estimators consistent? Using a theoretical analysis we could be able to answer these questions. This might lead to adaptations that improve the performance. If that is not possible, we could look into other ways to estimate the location of the change point. There might be other quantities present in the graph that can tell us something about the change point.

References

[1] Lada A. Adamic and Bernardo A. Huberman. “Power-Law Distribution of the World Wide Web”. In: *Science* 287.5461 (2000), pp. 2115–2115. DOI: 10.1126/science.287.5461.2115a. eprint: https://www.science.org/doi/pdf/10.1126/science.287.5461.2115a. URL: https://www.science.org/doi/abs/10.1126/science.287.5461.2115a.

[2] Simone Baldassarri and Gianmarco Bet. *Asymptotic normality of degree counts in a general preferential attachment model*. 2021. arXiv: 2102.07570 [math.PR].

[3] Pritha Bandari. *Type I & Type II Errors — Differences, Examples, Visualizations*. https://www.scribbr.com/statistics/type-i-and-type-ii-errors/. (Accessed on 05/24/2024). Jan. 2021.

[4] Albert-László Barabási. “Network science”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1987 (2013), p. 20120375.

[5] Albert-László Barabási and Réka Albert. “Emergence of Scaling in Random Networks”. In: *Science* 286.5439 (1999), pp. 509–512. DOI: 10.1126/science.286.5439.509. eprint: https://www.science.org/doi/pdf/10.1126/science.286.5439.509. URL: https://www.science.org/doi/abs/10.1126/science.286.5439.509.

[6] Gianmarco Bet, Kay Bogerd, Rui Castro, and Remco van der Hofstad. *Detecting a late changepoint in the preferential attachment model*. 2023. arXiv: 2310.02603 [math.ST].

[7] Shankar Bhamidi, Jimmy Jin, and Andrew Nobel. “Change point detection in network models: Preferential attachment and long range dependence”. In: *The Annals of Applied Probability* 28.1 (2018), pp. 35–78. ISSN: 10505164, 21688737. URL: https://www.jstor.org/stable/26542304 (visited on 09/18/2023).

[8] Kay Bogerd. “Detecting planted structures in random graphs”. English. Proefschrift. Phd Thesis 1 (Research TU/e / Graduation TU/e). Mathematics and Computer Science, Feb. 2021. ISBN: 978-90-386-5211-5.

[9] Béla Bollobás, Christian Borgs, Jennifer Chayes, and Oliver Riordan. “Directed scale-free graphs”. In: *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA ’03. Baltimore, Maryland: Society for Industrial and Applied Mathematics, 2003, pp. 132–139. ISBN: 0898715385.

[10] Béla Bollobás and Oliver Riordan. “The Diameter of a Scale-Free Random Graph”. In: *Combinatorica* 24.1 (Jan. 2004), pp. 5–34. URL: http://dx.doi.org/10.1007/s00493-004-0002-2.

[11] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. “The degree sequence of a scale-free random graph process”. In: *Random Structures & Algorithms* 18.3 (2001), pp. 279–290. DOI: https://doi.org/10.1002/rsa.1009. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.1009. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.1009.

[12] Daniel Cirkovic, Tiandong Wang, and Xianyang Zhang. *Likelihood-based Inference for Random Networks with Changepoints*. 2023. arXiv: 2206.01076 [stat.ME].

[13] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. “On Power-Law Relationships of the Internet Topology”. In: *ACM SIGCOMM Computer Communication Review* 29 (Mar. 2003). DOI: 10.1145/316188.316229.

[14] Fengnan Gao and Aad van der Vaart. “On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees”. In: *Stochastic Processes and their Applications* 127.11 (2017), pp. 3754–3775. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2017.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0304414917300819.

[15] Martijn Gösgens and Robert Fitzner. *Network Pages*. https://www.networkpages.nl/CustomMedia/Animations/RandomGraph/CM/PowerLaw_ShowDistance.html.

[16] Allan Gut. *An Intermediate Course in Probability*. Springer Texts in Statistics. Springer New York, 2009. ISBN: 9781441901620. URL: https://books.google.nl/books?id=ufxMwdtrm0AC.

[17] Oskar Hagberg and Carsten Wiuf. “Convergence properties of the degree distribution of some growing network models”. In: *Bulletin of Mathematical Biology* 68 (2006), pp. 1275–1291. URL: https://api.semanticscholar.org/CorpusID:8778127.

[18] Remco van der Hofstad. *Random Graphs and Complex Networks*. Vol. 1. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016. DOI: 10.1017/9781316779422.

[19] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan Oltvai, and Albert-László Barabási. “The large-scale organization of metabolic networks”. In: *Nature* 407.6804 (Oct. 2000), pp. 651–654. ISSN: 1476-4687. DOI: 10.1038/35036627. URL: http://dx.doi.org/10.1038/35036627.

[20] Miha Mozina. *Which is worse: False-positive or false-negative?* https://www.linkedin.com/pulse/which-worse-false-positive-false-negative-miha-mozina-phd. (Accessed on 05/19/2024). Nov. 2020.

[21] NASA. *Blackout Leaves American Cities in the Dark*. https://earthobservatory.nasa.gov/images/3719/blackout-leaves-american-cities-in-the-dark. (Accessed on 04/23/2024).

[22] Mark Newman. “The Structure of Scientific Collaboration Networks”. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.2 (2001), pp. 404–409. ISSN: 00278424. URL: http://www.jstor.org/stable/3054694 (visited on 12/18/2023).

[23] Jerzy Neyman and Egon Pearson. “On the problem of the most efficient tests of statistical hypotheses”. In: *Philosophical Transactions of the Royal Society (A)* 231 (1933), pp. 289–337.

[24] Matjaž Perc. “The Matthew effect in empirical data”. In: *Journal of The Royal Society Interface* 11.98 (Sept. 2014). ISSN: 1742-5662. DOI: 10.1098/rsif.2014.0378. URL: http://dx.doi.org/10.1098/rsif.2014.0378.

[25] Duncan Watts and Steven Strogatz. “Collective dynamics of ’small-world’ networks”. In: Dec. 2011. ISBN: 9781400841356. DOI: 10.1515/9781400841356.301.

A Discrepancy between papers

During this thesis, the asymptotic covariance matrix presented in [2] has been derived for three specific instances. During the derivation, it became clear that the theoretical values obtained in this derivation and the simulated values obtained using the simulation in Appendix B did not match. Furthermore, when looking at the variance of the number of vertices with minimal degree derived in [6], it became clear that one of the two papers was wrong. We were able to prove that [2] contained an error, which has been communicated with the authors. The authors of [2] released an updated version with the correct formula as of April 2024. Below we shortly explain how this error was uncovered.

To explain the discrepancy we are encountering, We fill in $l = m$ and $r = m$ into $R_Z(l, r)$ in the paper of [2] and show the different steps to obtain $R_Z(m, m)$. We start with the first part. The first problem to tackle is the $b_j^{(k)}$ terms. We notice that for $l, r = m$, the only 2 possibilities are

$$
\begin{aligned}
b_m^{(m)} &= 1, \\
b_q^{(m)} &= \begin{cases} 1 & \text{for } q = m, \\ 0 & \text{for } q > m. \end{cases}
\end{aligned}
$$

Therefore in the first term of the infinite sum, we only have 3 terms. These are the first, second and fourth line. We notice that as soon as $q = m + 1$, only the first term is still relevant. The rest all become zero due to $b_q^{(m)} = 0$ if $q = m + 1$. This gives us the following result

$$
\begin{aligned}
\frac{1}{\Gamma(4m + 3\delta_0 + 1)} \Bigg(&(m + \delta_0) p_m(\delta_0) \Big(\Gamma(4m + 3\delta_0) - 2\Gamma(4m + 3\delta_0) + m\Gamma(4m + 3\delta_0) \Big) \\
&+ \sum_{q=m+1}^{\infty} (q + \delta_0) p_q(\delta_0) \Gamma(4m + 3\delta_0) \Bigg).
\end{aligned}
$$

We take out the common term $\Gamma(4m + 3\delta_0)$. We can then immediately use the property of Gamma functions that states that

$$
\Gamma(x + 1) = x\Gamma(x).
$$

This gives us

$$
\begin{aligned}
&\frac{\Gamma(4m + 3\delta_0)}{\Gamma(4m + 3\delta_0 + 1)} \left((m + \delta_0) p_m(\delta_0) \Big(1 - 2 + m \Big) + \sum_{q=m+1}^{\infty} (q + \delta_0) p_q(\delta_0) \right) \\
&= \frac{1}{(4m + 3\delta_0)} \left((m + \delta_0) p_m(\delta_0)(m - 1) + \sum_{q=m+1}^{\infty} (q + \delta_0) p_q(\delta_0) \right).
\end{aligned}
$$

Now we add and subtract the missing term in the infinite sum, so we can solve it. We know 2 identities that can help solve the infinite sum

$$
\begin{aligned}
\sum_{q=m}^{\infty} p_q(\delta_0) &= 1, \\
\sum_{q=m}^{\infty} q p_q(\delta_0) &= 2m.
\end{aligned}
$$

We then get

$$\begin{aligned}
&\frac{1}{(4m+3\delta_0)}\Bigg((m+\delta_0)p_m(\delta_0)(m-1)-(m+\delta_0)p_m(\delta_0)+\sum_{q=m}^{\infty}(q+\delta_0)p_q(\delta_0)\Bigg)\\
&=\frac{1}{(4m+3\delta_0)}\Bigg((m+\delta_0)p_m(\delta_0)(m-2)+\sum_{q=m}^{\infty}qp_q(\delta_0)+\delta_0\sum_{q=m}^{\infty}p_q(\delta_0)\Bigg)\\
&=\frac{1}{(4m+3\delta_0)}\Bigg((m+\delta_0)p_m(\delta_0)(m-2)+2m+\delta_0\Bigg).
\end{aligned}$$

We finally plug in $p_m(\delta_0)=\frac{2+\frac{\delta_0}{m}}{m+2+\delta_0+\frac{\delta_0}{m}}$ and get

$$\frac{1}{(4m+3\delta_0)}\left(\frac{(m+\delta_0)(2+\frac{\delta_0}{m})(m-2)}{m+2+\delta_0+\frac{\delta_0}{m}}+2m+\delta_0\right).$$

For the second part of $R_Z(m,m)$, we notice that both double sums only have a single term. We also notice that the Gamma functions present in the second part cancel each other out, so we will omit them for efficiency. Moreover, every factorial in the denominator is zero, and $0!=1$, so we can omit those as well. Working out the sums, we get

$$-(2m+\delta_0)\Big(\frac{(4m+3\delta_0)^{-1}(m+1+\delta_0)^2}{(m+2+\delta_0+\frac{\delta_0}{m})^2}\Big)-\frac{(2m+\delta_0)(m-1)}{m^2}\Big(\frac{(m+\delta_0)^2(4m+3\delta_0)^{-1}}{(m+2+\delta_0+\frac{\delta_0}{m})^2}\Big).$$

Now we combine the two parts and work out the brackets. That gives us four specific terms:

$$\begin{aligned}
R_Z(m,m)=&\frac{(m+\delta_0)(2+\frac{\delta_0}{m})(m-2)}{(4m+3\delta_0)(m+2+\delta_0+\frac{\delta_0}{m})}+\frac{2m+\delta_0}{4m+3\delta_0}\\
&-\frac{(2m+\delta_0)(m+1+\delta_0)^2}{(4m+3\delta_0)(m+2+\delta_0+\frac{\delta_0}{m})^2}-\frac{(2m+\delta_0)(m-1)(m+\delta_0)^2}{m^2(4m+3\delta_0)(m+2+\delta_0+\frac{\delta_0}{m})^2}.
\end{aligned}$$

We can ask Mathematica to factorise these four terms together. By using the factor function in Mathematica on this expression, Mathematica gives us the following formula:

$$R_Z(m,m)=\frac{m^2(m+\delta_0)(m+\delta_0+1)(2m+\delta_0)}{(4m+3\delta_0)(m^2+m\delta_0+2m+\delta_0)^2}.$$

In the literature, the authors of [6] also derived the variance of the minimal degree in the preferential attachment model. They end up with the following result:

$$w(\delta_0,m)=\frac{m^2(m+\delta_0)(m+\delta_0+1)(2m+\delta_0)}{(\delta_0+2m(1+m+\delta_0))(m^2+m\delta_0+2m+\delta_0)^2}.$$

These terms seem very similar, except for one specific term in the denominator. However, this term matters a lot when $m>1$. For $m=1$ the different terms are identical however:

$$\begin{aligned}
\delta_0+2m(1+m+\delta_0)&=4m+3\delta_0\\
(2m+1)\delta_0+2m(m+1)&=4m+3\delta_0\\
(2m-2)\delta_0+2m(m-1)&=0\\
2(m-1)(m+\delta_0)&=0.
\end{aligned}$$

This means that for every simulation run with $m=1$, this difference will go unnoticed. However, these terms are not the same, and for $m>1$ there is a discrepancy between the result of [2] and [6]. As [6] has done extensive simulations to confirm the variance, this leads me to believe there is an error in the formula $R_Z(l,r)$.

B Pseudocode simulation

Algorithm 1 presented below is a degree sequence simulator. The algorithm is coded using Python. It does not simulate a whole graph with vertices and edges, but only simulates the degree sequence of a possible Preferential Attachment Model. After every new vertex, the algorithm updates the degree sequence. After the specified number of vertices n_0 is reached, the algorithm stops and gives the final degree sequence of the graph. In the definition of the code you can also specify $k \geq m$. This value will be the largest value that the algorithm keeps track of. If you specify $k = m + 1$ the algorithm only keeps track of the number of vertices with degree m and $m + 1$. It will also keep track of the sum of vertices that have degree $m + 2$ or larger, but these values are mainly for the algorithm to run correctly.

Algorithm 1 Degree sequence simulator

Require: n_0, m, $\delta_0 > -m$, $\delta_1 > -m$, τ, $k \geq m$

if $\delta_0 < -m$ or $\delta_1 < -m$ **then**
    Error: no valid input
Degree sequence $(N_m(n), N_{m+1}(n), \dots, N_k(n)) \leftarrow (2, 0, \dots, 0)$
$\delta(t) = \delta_0$
$n \leftarrow 2$
while n is less than n_0 **do**
    if n is equal to τ **then**
        $\delta(t) = \delta_1$
    for i in range of $\{1, \dots, m\}$ **do**
        Total degrees $\leftarrow (2m + \delta(t)) \cdot (n - 1) + i$
        Sample x from $X \sim Unif[0, 1]$
        for j in range of $\{m, \dots, k\}$ **do**
            $p \leftarrow (j \cdot N_j(n) + \delta(t)/\text{total degrees})$
            if $x < p$ **then**
                Pick $N_j(n)$ and break the loop.
            else $x \leftarrow (x - p)$
        Update degree sequence in the following way:
        $N_j(n) \leftarrow N_j(n) - 1.$
        $N_{j+1}(n) \leftarrow N_{j+1}(n) + 1.$
    $N_m(n) \leftarrow N_m(n) + 1.$
    Increase n by one
Return degree sequence $(N_m(n), N_{m+1}(n), \dots, N_k(n))$